Global analysis of cellular protein dynamics by pulse-labeling and quantitative mass spectrometry

D i s s e r t a t i o n

zur Erlangung des akademischen Grades

d o c t o r r e r u m n a t u r a l i u m

(Dr. rer. nat.)

im Fach Biologie

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät I

der Humboldt-Universität zu Berlin

von

Dipl.-Biol. Björn Schwanhäußer

Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Dr. h.c. Christoph Markschies

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I

Prof. Dr. Andreas Herrmann

Gutachter/innen: 1. Prof. Thomas Sommer

2. Prof. Matthias Selbach

3. Prof. Andreas Herrmann

Tag der mündlichen Prüfung: 23.02.2010

for Liam

Table of contents

Abstract

Gene expression is a tightly controlled process that is subject to transcriptional and as post-transcriptional regulation. Current methods for system-wide gene expression analysis detect changes in mRNA abundance but neglect regulation at the level of translation. The first part of the thesis therefore describes the establishment of modified version of the classic SILAC (stable isotope labeling by amino acids in cell culture) approach, that is routinely used in quantitative mass spectrometry to assay relative changes in protein levels. However, since steady-state protein levels reflect the net outcome of antagonizing protein synthesis and protein degradation processes, the SILAC method cannot be harnessed to explicitly measure differences in protein translation. In contrast, in the newly-devised approach termed pulsed SILAC (pSILAC) differentially treated cell populations are simultaneously transferred to culture medium supplemented with different versions of stable-isotope labeled heavy amino acids. This is advantageous over the classic SILAC strategy as mass spectrometry-based relative quantification is exclusively based on the newly-synthesized heavy protein amounts. This enables the specific detection of differences in protein translation resulting from the differential treatment. The second part of the thesis presents the application of the pSILAC approach to globally quantify the impact of small, non-coding RNAs (microRNAs) as major players in post-transcriptional regulation onto the proteome. Intriguingly, ectopic over-expression or knock-down of a single microRNA both affected protein production of hundreds of proteins. Notably, pSILAC identified several target genes as exclusively translationally regulated since changes in corresponding transcript levels measured in parallel by conventional DNA microarrays were virtually absent. Recording newly-synthesized protein amounts with heavy amino acids in a pulsed-labeling approach has also been used to determine turnover rates of individual proteins, described in the third part of the present work. Along with transcript turnover as well as mRNA and protein levels they form the basis for a dynamic description of gene expression. However, a systematic comparison of those parameters has not yet been performed in any organism. By simultaneous application of the nucleoside analogue 4-thiouridine (4sU) and heavy amino acids (SILAC) a metabolic labeling of newly-produced mRNAs and proteins in mouse fibroblasts has been achieved. This resulted in the calculation of mRNA and protein lifetimes for approximately 5,000 genes. On top of that, absolute cellular transcript and protein copy numbers were deduced from the exact same experimental data. While mRNA and protein levels were overall well correlated, a correlation between mRNA and protein half-lives was virtually absent. Yet this seemingly chaotic distribution of mRNA and protein half-lives was highly instructive since specific gene subsets have obviously evolved distinct combinations of half-lives that relate to their biological functions. For instance, dynamically regulated genes (e.g. transcription factors) were commonly characterized by both short mRNA and protein turnover what optimizes those gene products towards fast and flexible gene expression kinetics. As opposed to this, highly abundant and hence energetically costly gene products (e.g. structural proteins) had high mRNA and particularly high protein stabilities.

Zusammenfassung

Die Expression von Genen wird auf verschiedenen Stufen sehr präzise gesteuert, wobei grundsätzlich zwischen transkriptionaler und post-transkriptionaler Regulation zu unterscheiden ist. Mit den derzeit verfügbaren Methoden zur Genexpressionsanalyse können zwar relative Änderungen von mRNA-Mengen Mengen global erfasst werden, sie vernachlässigen jedoch den Einfluss translationaler Kontrolle. Der erste Teil der vorliegenden Arbeit beschreibt daher die Etablierung einer modifizierten Form des klassischen SILAC (engl. für „stable isotope labeling by amino acids in cell culture“) Verfahrens, das in der quantitativen Massenspektrometrie zur Bestimmung von relativen Änderungen in Proteinmengen benutzt wird. Allerdings kann dieser Ansatz nicht zur expliziten Messung von relativen Unterschieden in der Translation verwendet werden, da die gemessenen Abundanzunterschiede immer die Summe von Proteinneusynthese und Proteindegradation wiederspiegeln. Im sog. „pulsed SILAC (pSILAC)“ Verfahren werden Zellpopulationen im Zuge einer differentiellen Behandlung zeitgleich in Kulturmedien transferiert, die unterschiedlich Isotop-markierte und damit unterschiedlich „schwere“ Aminosäuren enthalten. Der entscheidende Unterschied gegenüber dem klassischen SILAC Prinzip besteht nun darin, dass die relative Quantifizierung im Massenspektrum ausschließlich auf dem Verhältnis der neusynthetisierten Proteinmengen beruht und so gezielt Unterschiede in der Produktion von Proteinen präzise bestimmt werden können. Mit Hilfe dieses Verfahrens konnte im zweiten Teil der Arbeit erstmals quantitativ erfasst werden, welche Rolle kleine regulatorische RNA Moleküle, sog. microRNAs, bei der post-transkriptionalen Genregulation spielen. So konnte gezeigt werden, dass sowohl die ektopische Überexpression als auch die Repression einzelner microRNAs die Proteinproduktion hunderter Proteine beeinflussen kann. Außerdem konnten mittels pSILAC Gene identifiziert werden, die ausschließlich translational reguliert werden, da parallele Messungen mit herkömmlichen DNA-Microarrays kaum Änderungen in den Transkriptmengen feststellen konnten. Die Messung von Proteinneusynthese mittels schwerer Aminosäuren ermöglichte auch die Bestimmung zellulärer Umsatzraten bzw. Halbwertszeiten individueller Proteine, dargestellt im dritten Teil dieser Arbeit. Zusammen mit mRNA-Halbwertszeiten sowie Protein- und mRNA-Mengen bilden sie die Grundlage für eine Beschreibung der dynamischen Eigenschaften der Genexpression, wobei eine systematische Analyse dieser vier Faktoren bisher noch nicht durchgeführt wurde. Durch den gleichzeitigen Einsatz des Nukleosidanalogons 4-Thiouridin (4sU) und von schweren Aminosäuren (SILAC) konnte eine metabolische Markierung neusynthetiserter mRNA und Proteine in murinen Fibroblasten erreicht und damit eine Berechnung von Protein- und mRNA Halbwertszeiten für ca. 5,000 Gene ermöglicht werden. Darüber hinaus konnten aus den gleichen experimentellen Daten zelluläre mRNA- und Proteinkopienzahlen abgeleitet werden. Während mRNA- und Proteinenmengen deutlich korrelierten, war zwischen mRNA- und Proteinhalbwertszeiten nur eine äußerste schwache Korrelation zu erkennen. Dennoch stehen mRNA- und Proteinumsatzraten nicht einem willkürlichen Zusammhang zu einander, da bestimmte Kombinationen von mRNA- und Proteinhalbwertszeiten eine Optimierung von Genen hinsichtlich ihrer biologischen Funktionen erkennen ließen. So zeichneten sich dynamisch regulierte Gene (z.b. Transkriptionsfaktoren) durch kurze mRNA- und Proteinhalbwertszeiten aus, während abundante und damit energetisch gesehen” teure” Genprodukte (z.b. Strukturproteine) hohe mRNA- und Proteinstabilitäten aufwiesen.

Abbreviations

3´UTR	3´ untranslated region
3D	Three dimensional
5´UTR	5´ untranslated region
ABC	Ammonium Bicarbonate
AHA	Azidohomoalanine
APEX	Absolute protein expression
AQUA	Absolute quantification
ATP	Adenosine triphosphate
BONCAT	Bioorthogonal non-canonical amino acid tagging
CDS	Coding sequence
cf.	Confer
CID	Collision induced dissociation
CMV	Cytomegalovirus
DDA	Data dependant acquisition
ddH_2O	Double distilled water
dFCS	Dialyzed fetal calf serum
DFO	Deferroxamine
DMEM	Dulbecco's Modified Eagle Medium
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dsRNA	Double stranded RNA
e.g.	Exempli gratia
emPAI	Exponentially modified protein abundance index
ER	Endoplasmic reticulum
ESI	Electrospray ionization
FAC	Ferric amonium citrate
FCS	Fetal calf serum
FDR	False discovery rate
FTH1	Ferritin heavy chain 1
FTL	Ferritin light chain
g	Gravity
GO	Gene Ontology
h	Hour(s)
i.e.	Id est (that is)
iBAQ	Intensity based absolute quantification
ICAT	Isotope-coded affinity tag
IRE	Iron responsive element
IRP	Iron regulatory protein
iTRAQ	Isobaric tag for relative and absolute quantification
kDa	Kilodalton
LC-MS	Liquid chromatography mass spectrometry

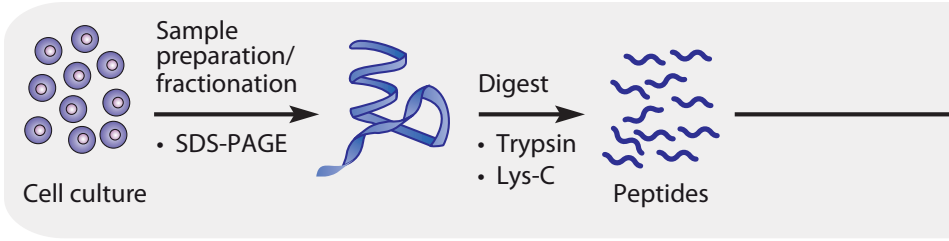
LNA	Locked nucleic acid
LTQ	Linear trap quadrupole
m/z	Mass-to-charge ratio
MALDI	Matrix assisted laser desorption/ionization
min	Minute(s)
miRNA	MicroRNA
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NanoLC	Nanoflow liquid chromatography
PAI	Protein Abundance Index
P-bodies	Processing bodies
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
ppb	Parts per billion
PPI	Protein-protein-interaction
ppm	Parts per million
pSILAC	pulsed SILAC
PTM	Post-translational modification
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
rpHPLC	Reversed-phase high performance liquid chromatography
RPKM	Reads per kilobase exon model per million mapped reads
rpm	Rounds per minute
R_s	Spearman´s rank correlation coefficient
R^2	squared Pearson correlation coefficient/ coefficient of determination
RT	Room temperature
s	Second(s)
SCX	Strong cation exchange
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SILAC	Stable isotope labeling by amino acid in cell culture
SEM	Standard error of the mean
SIM/MRM	Selected ion monitoring/multiple reaction monitoring
siRNA	Silencing RNA
TBS	Tris buffered saline
TCA	Tricarboxylic acid
TFA	Trifluoroacetic acid
TFRC	Transferrin receptor
TIC	Total ion current
uORF	Upstream open reading frame
V	Volt
XIC	Extracted ion chromatogram
YPD	Yeast peptone dextrose

I Introduction

With the rise of high-throughput technologies initially used to study entire genomes, referred to as genomics, the suffix "-omics" has been added to many fields in order to denote studies conducted on a global scale. Driven by the concept that studying biological molecules one by one cannot enable the understanding of function and dysfunction of whole organisms to a satisfactory degree, several disciplines have now evolved a multitude of technologies to study transcripts, proteins, metabolites and other biological entities on systems perspectives. Collectively, such approaches constitute the young yet exciting and highly interdisciplinary field of systems biology. The bird´s eye view of systems biology and its explicit goal to systematically integrate data from various disciplines may eventually allow scientists to comprehensively track signal transduction on different molecular layers throughout the cell. Ultimately, insights on a systems level will aid in closing the gap between single-molecule properties and the physiology of the entire cell (Bruggeman & Westerhoff, 2007).

Based on knowledge about genomic architecture gained by extensive genomic sequencing, researchers faced the challenge to annotate and characterize the plethora of proteins deriving from the strikingly smaller number of coding genes (Cravatt *et al.,* 2007). This obvious discrepancy arises from the fact that a single gene can give rise to multiple distinct proteins due to alternative splicing, sequence polymorphisms, post-translational modifications and so forth.

Despite the availability of exciting new technologies for rapid and paralleled analysis, including mass spectrometry (MS), protein microarrays, large scale two-hybrid-analyses, high throughput protein production and crystallization, the systematic analysis of all proteins in an organism, a tissue or a cell as the explicit goal of "proteomics" remains elusive so far (Wilkins, 1994; Aebersold & Mann, 2003). Due to the high degree of complexity of biological samples neither the proteome composition nor the quantity of its constituents can be reliably predicted by computational methods or determined by experimentation. The dynamic nature of proteomes that readily exceeds the dynamic range of any conventional analytical instrument poses hurdles much larger than encountered for genome or transcriptome studies (Mallick & Kuster, 2010).



Among diverse proteomics techniques, however, mass spectrometry has arguably become the dominant technology in the past decade for studying protein production and function in native biological systems. Contributions from cognate rapidly evolving disciplines like genomics and incremental but continuous instrumental innovations and improvements have paved the way for MS to become an indispensable tool for the emerging field of systems biology. Most importantly, the groundbreaking development of specific ionization techniques (particularly

ESI and MALDI, see below), awarded with the Nobel Prize in 2002, ultimately rendered biological molecules amenable to mass spectrometry and made instruments to become widespread in research laboratories (Cox & Mann, 2007). Due to the ever growing number of complete genomic sequences as a crucial resource for protein identification, sophisticated computational approaches and robust sample preparation methods, MS has now developed to the point at which it is routinely applied to address various research questions. At present, mass spectrometry based proteomics has become the method of choice to study differences in global protein compositions, for detection of post-translational modifications (PTMs) and protein-protein interactions (PPIs) as well as for protein biomarker discovery (Mallick & Kuster, 2010).

In the present thesis, mass spectrometry was used as the central technology to address various biological questions on a systems-level. As already apparent from the title of the thesis, those topics are all centered around cellular protein dynamics given by regulated protein translation and protein turnover. Since in particular protein translation could not be satisfactorily approached with currently available mass-spectrometry assays, the thesis therefore also has a methodological focus that describes the establishment of a method tailored to explicitly quantify differences in protein synthesis. For that reason, the following introduction will first describe the basic workflow and quantitative aspects of mass spectrometry and then discuss the biological context in which the technology was utilized to study cellular protein dynamics.

I.1. Mass spectrometry in a nutshell

I.1.1. Instrumentation and workflow of mass spectrometry

Modern mass spectrometers are able to accurately determine the mass (more precisely, the mass-to-charge ratio, m/z) of biological molecules. Knowing the exact mass of any given protein is, however, not sufficient for unique identification in a complex mixture as entirely different proteins may yet have identical masses. Moreover, protein handling, e.g. with regard

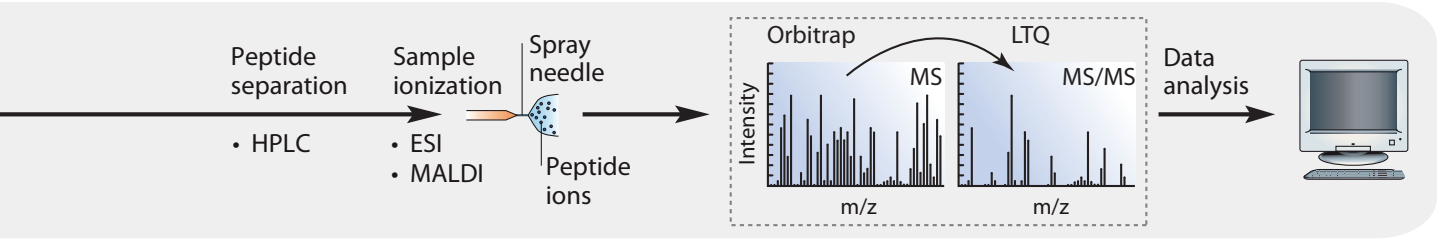


Fig. I.1 | Common workflow in mass spectrometry-based proteomics. Following sample preparation from a biological source (e.g. cell culture), proteins are subjected to enzymatic digestion. In order to decrease sample complexity, additional fractionation steps (e.g. SDS-PAGE) can optionally be performed before the digest (via trypsin or Lys-C). The generated peptide mixture is separated using reversed phase high performance liquid chromatography (rpHPLC) and directly subjected to mass spectrometry analysis via electrospray ionization (ESI). Alternatively, peptide ionization can be achieved by matrix-assisted laser desorption/ionization (MALDI). Given a LTQ-Orbitrap instrument, accurate mass determination of ionized peptides is performed in the Orbitrap part, while top N most abundant ions are concurrently selected and fragmented in the LTQ part to obtain sequence information. Finally, peptide-sequencing data deduced from acquired mass spectra are searched against sample specific protein databases for protein identification (Illustration taken and modified from Steen & Mann, 2004).

to protein solubility and ubiquitous protein modifications do further limit the capability to analyze intact proteins (STEEN & MANN, 2004). Hence, robust protein identification requires peptide information, and mass spectrometers work most efficient at peptide lengths up to 20 amino acids. Following protein isolation from any biological source (e.g. cell culture), proteins therefore usually undergo proteolytic digestion (cf. Fig. I.1). Although sample complexity even increases dramatically, the cleavage products are now much more amenable to mass spectrometry analysis. Digestion of most proteins produces at least some soluble peptides which also readily ionize and fragment, two requirements indispensable for effective mass spectrometry analysis. As peptide mixtures comprise thousands of peptides spanning several orders of magnitude in abundance, they cannot be satisfactorily analyzed all at a once. Therefore, separation techniques aiming at reduction of sample complexity have become key elements in proteomics experiments. One widely used and powerful setup is liquid chromatography mass spectrometry (LC-MS) where a high performance liquid chromatography (HPLC) system is coupled in-line with the mass spectrometer (STEEN & MANN, 2004). Of note, reversed-phase HPLC (rpHPLC) provides the highest resolving power for peptide mixtures currently available in proteomics. In rpHPLC, following injection of peptides onto microscale capillary columns filled with reversed-phase material, peptides sequentially elute in order of their hydrophobicity when a gradient with an increasing fraction of organic solvent is applied.

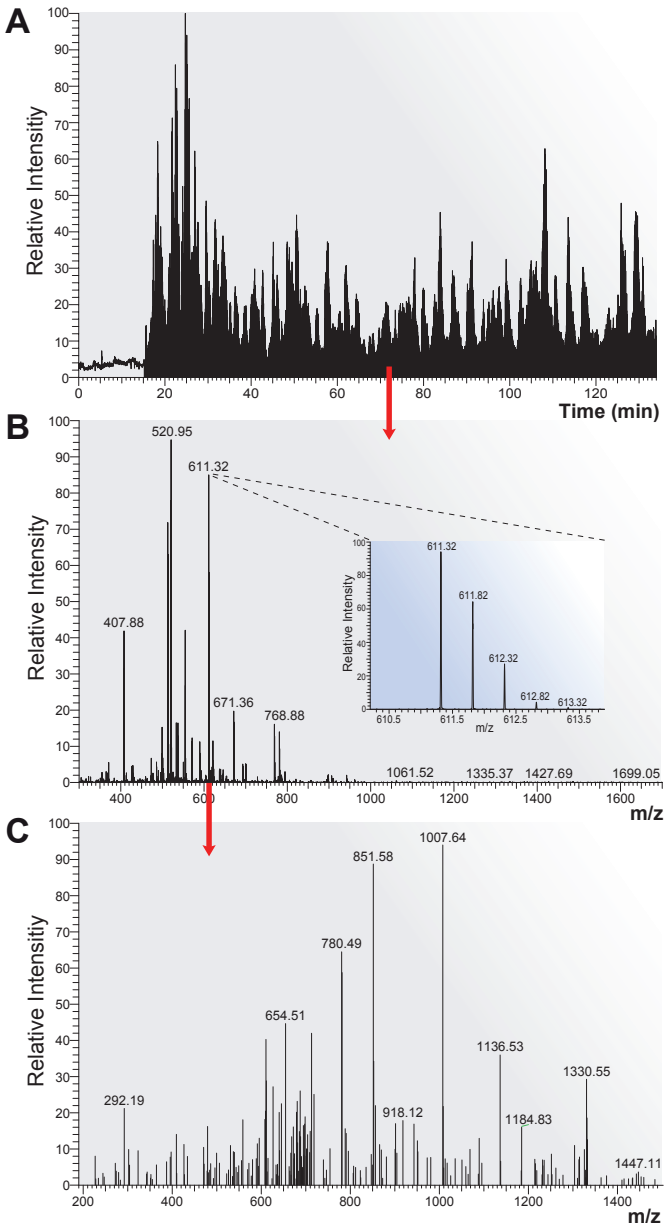


Fig. I.2 | Tandem mass spectrometry (MS/MS). (A) Total ion current (TIC) that is the sum of all ion intensities from all mass spectra being recorded during the LC-MS run as a function of time. (B) Survey/ precursor-ion scan depicting peptide ions within a m/z range of 300 - 1,700 arising from peptides eluting at a certain time point (here at 72.2 min). The insert shows the isotopic distribution (isotope cluster) of a given peptide ion (m/z of 611.32). A clear separation of individual isotope masses is indicative of a high resolution instrument. (C) Fragment or tandem MS spectrum of the peptide ion of interest (m/z of 611.32). Distinct mass increments between individual peaks allow for deduction of the amino acid sequence.

As mass spectrometric measurements are carried out in the gas phase on ionized molecules, peptides eluting from the column are electrostatically dispersed by means of high voltage applied to the column that generates positively charged peptide ions. This process is referred to as electrospray ionization (ESI) which rendered in-line coupling

of HPLC systems with mass spectrometers possible in the first place. Another widespread used ionization techniques is MALDI (matrix-assisted laser desorption/ionization), where a laser beam triggers ionization of biomolecules embedded in a chemical matrix. ESI and MALDI are both soft ionization techniques that enable the ionization of fragile biological molecules such as peptides without disrupting them. Upon ionization, peptide ions enter the instrument via transfer capillary and can be guided and manipulated by applying electric fields for further analysis.

By definition, a typical mass spectrometer is composed of three basic components that are an ion source, a mass analyzer and an ion detector (Guerrera & Kleiner, 2005). As all described experiments in the present thesis were exclusively analyzed using a LTQ-Orbitrap system, all further remarks refer to that type of instrument. In brief, the LTQ-Orbitrap instrument is a latest generation hybrid mass spectrometer combining the strengths of two conceptually distinct mass analyzers in one device. To be exact, the LTQ-Orbitrap combines a linear trap quadropole (LTQ) and an Orbitrap mass analyzer (Scigelova & Makarov, 2006).

ESI produces positively charged molecules that are first trapped in an electrostatic field moving around a central electrode referred to as orbital trap, or Orbitrap. To avoid unfavorable collision with any other form of matter, mass spectrometers operate in high-vacuum atmospheres. The mass analyzer now uses individual ion oscillations from peptides with different masses to accurately compute respective m/z values (inferred from oscillation frequencies) and signal intensities (inferred from oscillation amplitudes). When all the signal intensities being detected across the entire mass range are summed up and plotted against the time given by the length of the HPLC gradient, the total ion current (TIC) is obtained (cf. Fig. I.2 A). Looking specifically at all masses present in the Orbitrap at a specific point in time yields the so called survey or precursor ion scan (cf. Fig. I.2 B). Since a mass spectrometer separates and detects ions of slightly different masses, high mass resolution instruments easily discern different isotopes of a given peptide. Hence, ionized peptides do usually not appear as single, isolated peaks but as so called isotope clusters reflecting the natural $^{12}C/^{13}C$ ratio (cf. Fig. I.2 B insert) (Canas *et al.,* 2006).

In order to unambiguously discern peptides of the same mass but different amino acid order, sequence information is a necessity. This is achieved by employing a strategy termed tandem mass spectrometry (MS/MS) that involves breaking of isolated peptide ions into smaller fragments by collision with inert gas molecules (termed collision induced dissociation, CID). The process of recording such fragment mass spectra (cf Fig. I.2 C) is referred to as fragment or product ion scanning (Domon & Aebersold, 2006). The resulting randomly occurring peptide breaks produce fragment ions of specific m/z values from which the amino acid sequence can be deduced. In practice, tandem mass spectrometry is performed in the LTQ-part of the instrument. While a precursor ion spectrum is acquired in the Orbitrap, top N most abundant precursor ions are concurrently selected for CID fragmentation in the LTQ (cf. Fig. I.2.C), a process called data-dependant analysis (DDA). Spatially and functionally separating the acquisition of precursor and product ion spectra assures that the unique analytical capabilities

of either part of the LTQ-Orbitrap instrument are fully exploited. In mass spectrometry-based proteomics, several key parameters have been used to describe the performance characteristics of an instrument such as sensitivity, speed, dynamic range and resolution. The sensitivity of a mass spectrometer is determined by the smallest quantity of an analyte that can be detected, which is in the sub-femtomol range for peptides (corresponds to protein copies/cell < 100). The instrument´s speed is used to assess the number of spectra that can be acquired per unit time and basically depends on the cycle time (the time needed to go through a cascade of one survey scan (MS spectrum) followed by multiple product ions scans (MS/MS spectra)). While the dynamic range of an instrument denotes the range between the highest and lowest signal of an analyte detected in a single analysis, the resolution indicates the ability to distinguish peptide signals of slightly different m/z values (cf. Fig. I.2 B). As sensitivity and speed are crucial parameters to obtain thousands of instructive fragment spectra in a single run, CID is performed in the LTQ-part of the instrument whereas the acquisition of full scans requiring a high dynamic range and a high resolution is carried out in the Orbitrap-part.

Having acquired individual peptide mass and sequence information, a “probability-based” matching algorithm is applied that compares the acquired spectra to theoretical spectra obtained by *in silico* digestion and fragmentation of candidate peptides. Once a tandem mass spectrum is assigned a peptide sequence, a protein database is searched to identify the proteins where the peptide sequence is part of.

Collectively, the aforementioned tryptic digest of proteins followed by tandem MS in order to obtain peptide mass and sequence information form the central concept of what is commonly termed shotgun, bottom-up or peptide-centric proteomics. Shotgun proteomics is one of the most powerful and widespread approaches for characterization of complex protein samples so far (Marcotte, 2007). At present, if extensive sample pre-fraction and exhaustive precursor ion sequencing is applied, samples can be analyzed to an impressive depths of more than 6,000 proteins. Due to its unparalleled high sampling depth, shotgun proteomics is often said to be ideally suited for “open discovery experiments”, where even novel proteins may potentially be detected (Domon & Aebersold, 2010).

Peptide-centric proteomics does, however, have intrinsic limitations that should always kept in mind when conducting the actual measurement and interpretation of the obtained data afterwards. First, as the selection of precursor ions for fragmentation is done in an automatic fashion (DDA), it is inherently biased towards more abundant peptide species displaying proper signal-to-noise ratios. Minimizing the resulting systematic under-sampling of highly interesting but low abundant proteins is therefore a major challenge towards a holistic proteome analysis (States *et al.*, 2006). Second, since signals intensities of the very same peptide may substantially differ between runs due to technical constraints (cf. I.1.2), DDA-based mass spectrometry results are intrinsically to some degree irreproducible. Consequently, different subsets of the proteome are identified after repeated analysis of identical samples and the apparent absence of a certain peptide cannot be construed as evidence for the absence of the corresponding protein. Identification of peptides may simply fail due to (i) being present at very low amounts

only, (ii) fragment spectra of poor quality, (iii) absence of the sequence from the database or (iv) extensive post-translational modifications that aggravates peptide identification (Duncan *et al.,* 2010).

Third, although tremendous effort has been devoted to optimize computational algorithms that match the measured fragment spectra to those obtained by *in silico* fragmentation, protein assembly in shotgun proteomics remains a non-trivial, probability-based central challenge to robust protein identification. Although the use of sequence-specific proteases along with profound mechanistic understanding on how and where peptides break apart vastly facilitates the assembly of proteins thereafter, the process is oftentimes compared to a jigsaw puzzle with billions of parts (Marcotte, 2003). Correct peptide assignment is further complicated as precedent protein digestion results in the loss of intact protein information. This applies especially to degenerated peptides, i.e. peptides shared by multiple proteins, where the enzymatic cleavage renders a definite precursor finding impossible. Although the bulk spectra can be eventually correctly assigned to peptide sequences, an appreciable amount of fragment ions go unmatched and remain unnoticed. Amongst them, spectra of multiply modified peptides but also spectra from novel peptides are assumed (Ning *et al.*, 2010).

Lastly, conducting shotgun proteomics virtually always means perpetual *de novo* discovery of a proteome, as the analysis can inherently not be restricted to a certain subset of proteins. Consequently, a large number of proteins with no relevance to the actual question are identified making peptide-centric proteomics a relatively highly time-consuming and costly process. Thus, if only a well defined set of proteins is of interest so called “targeted” MS approaches constitute reasonable alternatives (Domon & Aebersold, 2010). This being said, such approaches, however, require comprehensive *a priori* knowledge for both target selection and assay design - knowledge that is normally gained in upfront shotgun proteomics.

I.1.2. Quantitative approaches in mass spectrometry-based proteomics

I.1.2.1 Relative MS-based quantification with heavy stable-isotopes

Robust protein identification has become straightforward, yet various pressing questions cannot be reasonably addressed without additional quantitative information. Current improvements in genomic sequencing technologies, though imperfect, made whole transcriptomes accessible to quantitative descriptions (Schuster , 2008; Wang *et al.,* 2009). Unfortunately, mRNA abundances show only poor correlation with corresponding protein levels (Lu *et al.,* 2007; Gygi *et al.,* 1999; Greenbaum *et al.,* 2003). Thus, the ability to monitor changes in protein abundance, preferably on a global scale, is a key prerequisite for interpretation of virtually all targeted perturbations of biological systems. Moreover, any kind of modeling study taking dynamic data as input is inconceivable without a quantitative dimension (Hanke *et al.*, 2008).

Unfortunately, mass spectrometry inherently is a non-quantitative method. Specifically, the

raw peptide signal intensity is not directly proportional to the quantity of the peptide present in the sample due to multiple error-prone processing steps in MS. Differential accessibility of proteins to respective proteases, varying peptide solubility, differences in detector sensitivity for certain peptides and, most importantly, inconsistent ionization efficiencies do all affect raw peptide signal intensities (Ong & Mann, 2005). Yet the systematic character of those factors allows for relative comparison of ion signals, either between two samples analyzed in the same manner or within the same analysis by means of stable-isotopes (Aebersold & Mann, 2003).

In recent years, several methods utilizing stable isotopes for mass spectrometry- based quantification have been successfully established. Though differing in the way stable-isotope labeling is accomplished, all approaches share the key principle that pairs of chemically equivalent peptides of different stable-isotope composition can be distinguished by the mass spectrometer owing to the mass shift introduced by the label (Bantscheff *et al.,* 2007). Provided that the instrument is capable to detect and differentiate both the labeled and the unlabeled form, quantification is achieved by comparing the respective signal intensities. If thoroughly performed, stable-isotope labeling is characterized by outstanding quantification accuracy since comparison of signal intensities of the two peptide forms is performed within the same LC-MS run. Thus, differences in sample preparation and most notably in ionization efficiencies between runs do not apply here and therefore fundamentally distinguish isotope-labeling from so called “label-free” approaches mentioned thereinafter.

Stable-isotope labels are usually introduced by chemical tagging of peptides (termed ICAT for isotope-coded affinity tag introduced by Gygi *et al.,* 1999) or proteins (referred to as iTRAQ for isobaric tag described in Ross *et al.*, 2004) and also by transfer of heavy oxygen (^{18}O) during enzymatic protein digestion (Yao *et al.,* 2001). Apart from chemical and enzymatic labeling strategies, metabolic incorporation of heavy stable-isotopes is frequently used (cf. Fig.I.3). In essence, metabolic labeling is based on supplementing growth media of living cells with heavy stable-isotope molecules (Ong & Mann, 2005; Aebersold & Mann, 2003, Bantscheff *et al.,* 2007). Through cellular growth

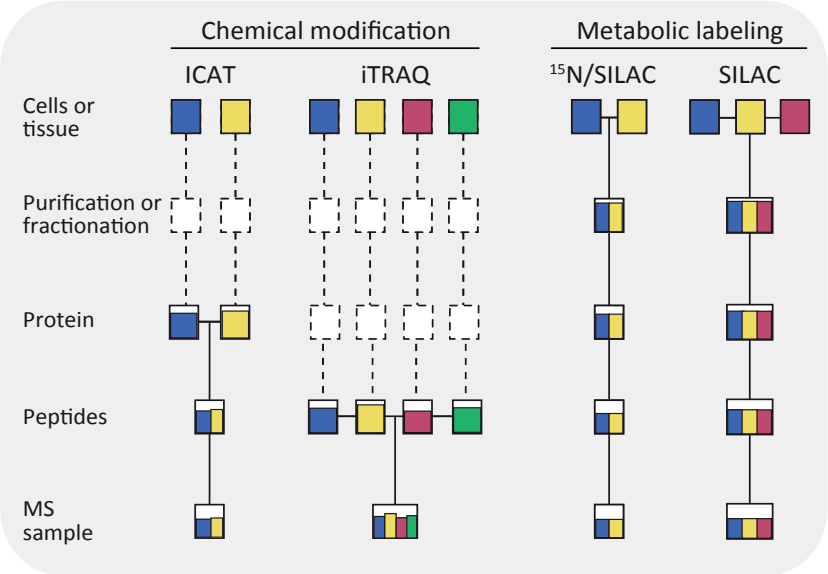


Fig. I.3 | Quantitative mass spectrometry workflows employing heavy stable-isotopes. Colors represent different experimental conditions being compared. Depending on the approach used to introduce stable-isotopes samples are combined at different steps during sample preparation, indicated by horizontal lines between boxes. Latest techniques enable multiplexing of samples, that is, comparing up to four samples in iTRAQ and up to three protein populations in stable isotope labeling by amino acids in cell culture (SILAC). While metabolically labeled samples (^{15}N/SILAC) are combined at the earliest possible step, that is, at the stage of intact cells, parallel sample processing (indicated by dashed lines/boxes) in chemical labeling methods such as ICAT and iTRAQ before mixing is potentially error-prone. (Scheme taken from Ong & Mann, 2007)

and protein turnover labeled metabolic precursors are gradually incorporated into virtually all proteins. Accordingly, cells are either grown in light (unlabeled) or heavy (labeled) growth media resulting in two different cell/protein populations distinguishable only by MS due to the distinct mass shift. As labeling takes place in living cells and therefore at the earliest possible stage, the samples of interest can be combined prior to lysis, fractionation and purification. This implies that errors introduced by the aforementioned sample preparation steps cancel out because they apply equally to both samples - a unique feature of metabolic labeling (cf. Fig. I.3).

Metabolic labeling can be distinguished in two major approaches given by ^{15}N labeling and stable isotope labeling by amino acids in cell culture (SILAC). Since all relative quantification experiments described in the present thesis were exclusively performed using SILAC, further specifications of chemical, enzymatic and ^{15}N-labeling assays are beyond the scope of the thesis and shall remain unnoticed. Strengths and weaknesses, limitations and recent applications of the SILAC approach, however, will be discussed in greater detail in the following chapter.

I.1.2.2 Stable isotope labeling by amino acids in cell cultue (SILAC)

In recent years, a straightforward yet simple and widely applicable metabolic labeling strategy termed SILAC (stable isotope labeling by amino acids in cell culture) has gained wide popularity. SILAC was originally developed in the lab of Matthias Mann in 2002 and involves the growth of cells in culture medium supplemented with heavy stable isotope versions of essential amino acids like heavy arginine and/or heavy lysine containing a fixed number of ^{13}C, ^{15}N and/or ^{2}H atoms (Ong *et al.,* 2002). In SILAC, potentially all peptides are amenable to labeling as the amino acids are uniformly incorporated into all newly synthesized proteins. Even proteins with extremely low turnover rates will be labeled due to dilution of unlabeled counterparts after several rounds of cell doubling (Ong & Mann, 2006). Thus, a typical SILAC experiment involves cultivation of two different cell populations in either light medium containing amino acids with natural isotopes or heavy medium containing the respective heavy amino acids (cf. Fig. I.4). In contrast to radioactive labeling commonly used in biology (Pollard, 1996), such as pulse–chase experiments with ^{32}P- or ^{35}S-Met, SILAC seeks to completely replace light amino acids by their heavy labeled counterparts throughout the proteome. As mentioned above, sample combination right after the actual experiment minimizes errors introduced by sample processing. As labeling does not affect chromatographic characteristics, co-eluting pairs of chemically equivalent peptides emerge in the mass spectrum at the same time but can still be distinguished from each other due to the constant mass increment introduced by heavy amino acids. Hence, comparing signal intensities of heavy and light peptide peaks allows for accurate relative quantification of thousands of proteins in one LC-MS/MS experiment (Ong & Mann, 2005).

Combining SILAC labeling with tryptic cleavage occurring specifically at the carboxyl-termini of arginine and lysine residues typically generates peptides carrying only one labeled amino acid which significantly facilitates subsequent data analysis. Knowing the exact mass difference

introduced during metabolic labeling is a fundamental difference to ^{15}N labeling where the number of atoms incorporated can vary depending on the peptide sequence (Ong & Mann, 2006). Also, proteolytic cleavage after heavy amino acids basically enables quantification of all peptides except for the C-terminal peptide of a protein.

Depending on the label used in SILAC, light and heavy peptide peaks differ in mass from at least four up to ten Dalton. This is usually sufficient to allow for a clear-cut discrimination between the respective isotope clusters, which is a major prerequisite for accurate quantification. Although SILAC works most convenient with immortalized cell lines where usually ~97 % of all proteins are present in the heavy state after five cell doublings in appropriate growth medium (Ong & Mann, 2006), recent studies demonstrate that the labeling approach can also be expanded to higher organisms. Despite of high costs and laborious experimental work, *in vivo* SILAC labeling of *Mus musculus* was achieved in 2008 (Kruger *et al.,* 2008). The "SILAC mouse" allowed for the first time ever organ and tissue specific proteome-wide relative quantification and represents undoubtedly a milestone in stable-isotope labeling. Alternatively, SILAC labeling of *Drosophila melanogaster* as an animal model amenable to straightforward genetic manipulation has also been accomplished (Sury *et al.,* 2010). The approach is simple, fast, and cost-effective which makes SILAC flies an attractive model system for the emerging field of *in vivo* quantitative proteomics.

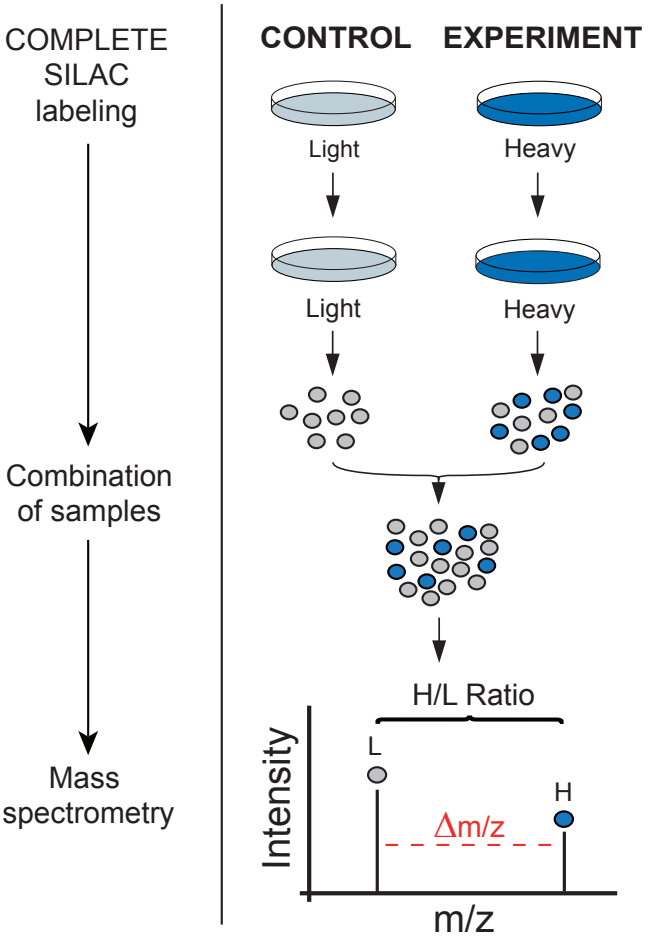


Fig. I.4 | Quantitative mass spectrometry using the SILAC approach. Cells of two different experimental conditions are either grown in culture medium containing unlabeled, natural amino acids (light medium) or in medium supplemented with heavy stable isotope versions of essential amino acids (heavy medium). As heavy amino acids $^{13}C_6{}^{15}N_4$ L-arginine (mass shift of ten Dalton to its unlabeled counterpart) and $^{13}C_6{}^{15}N_2$ L-lysine (mass shift of eight Dalton to its unlabeled counterpart) are preferably used. Cell growth, i.e. protein synthesis and turnover lead to uniform metabolic incorporation of the SILAC amino acids throughout the entire proteome. Intact cells from the two populations being compared are combined, processed, and subjected to mass spectrometry analysis. Here, pairs of chemically equivalent peptides emerge with a distinct mass shift (Δm/z) introduced by the labeling in the same mass spectrum. Comparing signal intensities of heavy and light peptide peaks allows for accurate relative quantification of thousands of proteins in one LC-MS/MS analysis.

Despite its appealing simplicity, SILAC does of course have limitations. For instance, the use of fetal calf serum (FCS) in cell culture is a potential source of unlabeled amino acids. Therefore, dialyzed FCS (dFCS) has to be used that is, however, incompatible with some rare cell lines relying on growth factors lost by dialysis. Moreover, certain cell types exhibit some degree of arginine to proline conversion. Fortunately, the conversion can be alleviated by titrating the amount of arginine used in the media (Ong & Mann, 2006). Lastly, cell lines

with extremely slow cell division rates (e.g. Schwann cells) can make a proper labeling highly time-consuming. As all of the aforementioned issues potentially affect proper SILAC labeling and hence quantification accuracy, they should be explicitly addressed before conducting the actual experiments.

Due to its straightforward implementation in existing workflows and its robust and highly accurate quantification of thousands of proteins in parallel, SILAC has become the method of choice for addressing various research subjects - from whole proteome scale comparisons (Graumann *et al.,* 2008) to acquisition of temporal profiles (Andersen *et al.,* 2005) and analysis of changes in post-translational modifications like phosphorylations (Blagoev *et al.,* 2004).

I.1.2.3 Absolute quantification of proteins

Stable-isotope based quantification has so far solely been discussed with regard to comparison of perturbed biological samples relative to an untreated control sample. Relative quantification, however, cannot provide information about absolute protein abundance. Especially in a medical context, knowing absolute amounts of disease-specifc biomarkers can provide diagnostic information of high relevance (Hanke *et al.,* 2008). Also, modeling approaches require absolute molecule numbers to quantitatively describe dynamic systems (Ben-Tabou de-Leon & Davidson, 2009).

Stable-isotope labeling can also be employed to obtain absolute protein quantification. Historically, digestion in heavy water was the first method used for absolute quantification (Desiderio & Kai, 1983). More recent approaches employ spiking of known amounts of a stable-isotope labeled peptides or an entire protein as an internal standard to the protein mixture, two concepts termed “AQUA” (Gerber *et al.,* 2003) or “Absolute SILAC” (Hanke *et al.,* 2008), respectively. Despite of its high precision in quantification, laborious and costly stable-isotope peptide/protein synthesis limits the approach so far to small scale protein studies. Alternatively, absolute quantification is often performed without heavy stable isotopes in so called “label-free” approaches, which are briefly discussed in the following chapter.

I.1.2.4 Label-free approaches

Conceptually different to isotope labeling are so called “label-free” approaches for relative and absolute quantification. Most methods rely on the observation that more abundant proteins are also more likely to be detected at the peptide level in a shotgun experiment (Liu *et al.,* 2004). These approaches either involve counting of fragment spectra obtained for a given protein (referred to as spectral counting) or measuring MS signal intensities of peptides derived from the same protein (referred to as extracted ion current, XIC). Spectral counting-based absolute quantification provides an estimate of absolute protein amounts, especially when corrected for the number of theoretically observable peptides (exponentially modified protein abundance index, emPAI) (Ishihama *et al.*, 2005) or sequence length (normalized spectral abundance factor, NSAF) (Mosley *et al.,* 2009). Combined with complex computational algorithms predicting the probability of identifying peptides using machine learning techniques (absolute

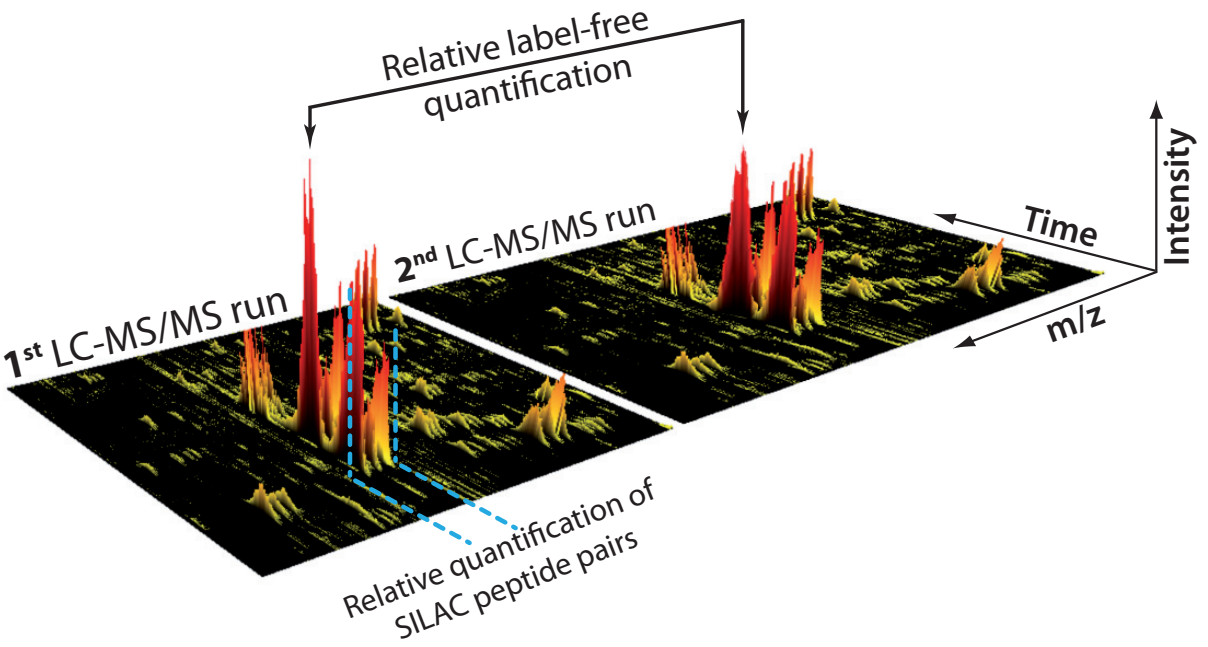


Fig. I.5 | Strategies used in quantitative mass spectrometry. While isotope labeling approaches such as SILAC compare peak intensities of differentially labeled peptide pairs with a specific mass shift in the same LC-MS/MS run, label-free methods for relative quantification of protein abundance are performed across runs.

protein expression measurements, APEX) (Lu *et al.,* 2007), spectral counting has gained more precision in quantification. Still, a limitation of spectral counting-based approaches is that they fully ignore peptide intensities as a possible source of quantitative information. Alternatively, an XIC-based approach taking into account raw intensities of the three most intense peptides matching to a specific protein has been demonstrated to correlate with absolute levels (Silva *et al.,* 2006; Malmstrom *et al.,* 2009). However, this approach neglects information from intensities of all other matching peptides and yields a smaller number of quantifiable proteins because the identification of three peptides per protein cannot be guaranteed in all cases.

Relative quantification employing label-free methods is further complicated by the fact that the comparison of peptide abundance is performed across LC-MS runs, whereas signal intensities of stable-isotope labeled peptides are directly compared within a single run (cf. Fig. I.5). Thus, differences in ionization efficiency and more important, slightly varying chromatographic retention times between runs may dramatically affect relative quantification in label-free approaches.

I.2. Biological implications of mass spectrometry-based quantitative proteomics

I.2.1. Translational regulation of gene expression

Gene expression is tightly regulated at all levels from mRNA transcription to protein translation and hence transcriptional and post-transcriptional regulation has to be distinguished. While the former one is crucial for mRNA synthesis, post-transcriptional control affects mRNA stability as well as distinct steps during translation (Gebauer & Hentze, 2004). In the past decades, much effort has been invested to study transcription regulation, while translational regulation largely remained unnoticed.

Comparative genomic and proteomic profiling conducted in different species have revealed a substantial lack of correlation between mRNA and protein levels thereby clearly stressing the relevance of post-transcriptional regulation for gene expression (Gygi *et al.*, 1999; Greenbaum *et al.,* 2003; Lu *et al.,* 2007). This observation implies that DNA-microarrays which have become a powerful and straightforward tool to routinely study quantitative changes of the transcriptome cannot be employed to study changes occurring at the level of translation.

Translational regulation provides the cell with the plasticity needed to rapidly adapt gene expression to changes in environmental conditions such as cellular stress (e.g. heat shock, hypoxia or nutrient deprivation), but also for fine-tuning of protein levels during cell cycle, proliferation and differentiation (Holcik & Sonenberg, 2005; Calkhoven *et al.,* 2002). In the cell, translational control is either exerted globally, thereby affecting basically all mRNAs or more selectively, in which translation of merely defined groups of mRNAs is modulated without affecting overall biosynthesis (Gebauer & Hentze, 2004). While global control of translation is primarily mediated by modifying translation initiation itself, specifically modulated mRNAs frequently carry *cis*-regulatory sequence motifs targeted by RNA-binding proteins upon environmental changes (Lunde *et al.,* 2007). Apart from proteins a class of small RNAs referred to as microRNAs (miRNAs, see below) has also been found to have a major impact on translational regulation by hybridizing to specific sequences usually located in untranslated regions (UTRs) (Flynt & Lai, 2008). Translation of localized mRNAs via functionallly distinct compartments represents another powerful means of translational control (Besse & Ephrussi, 2008). Consistently, a growing catalogue of human diseases is linked to mutations in gene products that are either part of the translation machinery or directly controlling translational regulation (Scheper *et al.,* 2007).

I.2.2. Regulation of gene expression exerted by microRNAs

Since their discovery in the early 1990, intense research on gene regulation exerted by microRNAs (miRNAs) has undoubtedly added a new dimension to the complexity of post-transcriptional regulation. As things turned out soon, this class of small non-coding RNAs

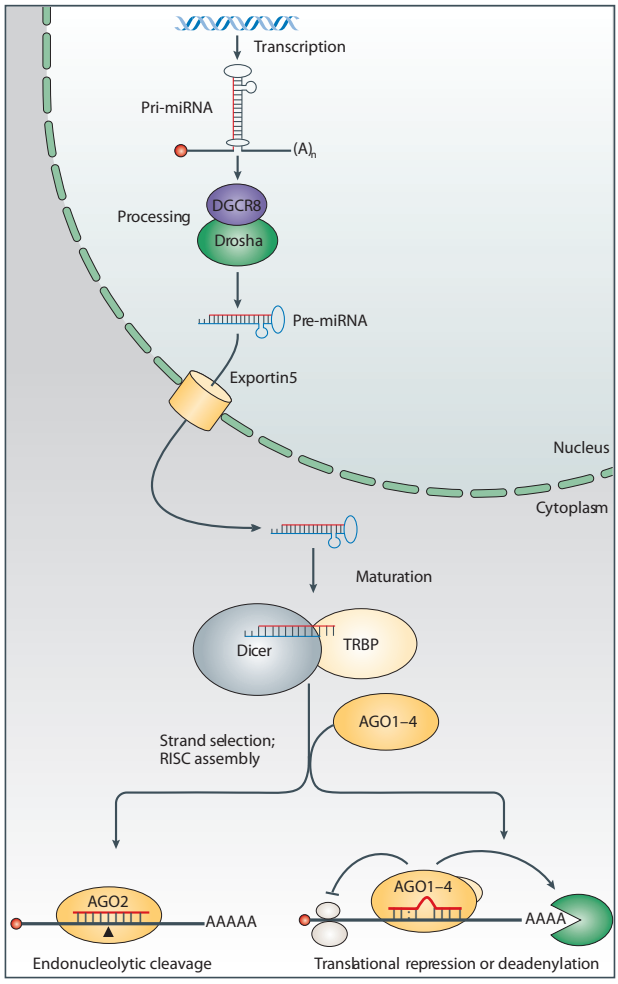


Fig. I.6 | MicroRNA biogenesis and putative mode of action in mammals. Following transcription of microRNA genes, the transcripts undergo a series of endonucleolytic maturation steps. First, primary miRNA transcripts (pri-miRNA) are processed into precursor miRNA (pre-miRNAs) stem loops by the RNaseIII enzyme Drosha and the double-stranded RNA binding-domain protein DiGeorge syndrome critical region gene 8 (DGCR8). Pre-miRNAs are then actively transported to the cytoplasm via Exportin 5 and cleaved into short double-stranded RNA duplexes by Dicer (complexed with TAR RNA binding protein (TRBP)). Only one strand (guide strand) of the duplex is selected and assembled together with specific Argonaute proteins (Ago 1-4) into the RNA-induced silencing complex (RISC). Finally, RISC represses target gene expression by either interfering with translation initiation/elongation or by deadenylation followed by destabilizing of target mRNAs. RISC also mediates RNA interference (RNAi) where full complementarity between small interfering RNAs (siRNAs) and target mRNAs results in endonucleolytic cleavage thereof. (Illustration taken and modified from FILIPOWICZ *et al.*, 2008)

was more than a mere curiosity of nematode biology (LEE *et al.,* 1993; WIGHTMAN *et al.*, 1993). Due to their vital role in virtually every biological process under investigation, aberrant expression levels are nowadays observed in a growing number of human pathologies, including cancer and neurodegenerative disorders (ESQUELA-KERSCHER & SLACK, 2006; KOSIK, 2006; CHANG & MENDELL, 2007). Despite of their ubiquitous nature, progress towards an exhaustive mechanistic understanding by which miRNAs act post-transcriptionally has been difficult as results from studies conducted in different systems and different laboratories have often been contradictory.

Metazoan microRNAs, typically 21-23 nucleotides in length, base-pair imperfectly with target mRNAs to down-regulate gene expression by either destabilization of the transcript or by translational repression (cf. Fig. I.6) (HE & HANNON, 2004; BRODERSEN & VOINNET, 2009). The most stringent requirement for target selection is a consecutive and perfect Watson-Crick base pairing of the miRNA nucleotides 2–8, representing the "seed" region, with one or multiple complementary *cis*-regulatory sites in the 3´UTR of target messages (STARK *et al.,* 2003; RAJEWSKY & SOCCI, 2004). While many target mRNAs undergo degradation, most likely initiated by removal of the poly(A) tail, other, stable target mRNAs are obviously repressed at the initiation or elongation stage of translation (EULALIO *et al.,* 2008; FILIPOWICZ *et al.,* 2008).

Bioinformatic predictions accounting for the above mentioned miRNA-mRNA binding characteristics indicate that mammalian miRNAs potentially regulate up to 30% of all protein-coding genes (RAJEWSKY, 2006). However, of the hundreds of known human miRNAs only very few have been experimentally linked to specific functions. The fact that microRNAs can alter mRNA stability has been exploited to study the impact of microRNAs following ectopic over-expression on global mRNA levels by conventional microarray technique (LIM *et al.*, 2005).

In contrast, experimental evidence demonstrating that microRNAs can also directly affect protein translation is restricted to a small set of genes (Mathonet *et al.,* 2007; Thermann & Hentze, 2007; Wakiyama *et al.,* 2007). Evidently, these results do not reveal how much control miRNAs exert globally on protein synthesis. Given that protein synthesis is one of the most important quantities for the phenotype, fundamental questions about gene regulation mediated by microRNAs therefore remain unanswered. Likewise, it is unclear whether sequence determinants beyond seed pairing can explain differences of individual microRNAs in targeting mRNAs for degradation or translational repression.

Recently, ribosome profiling has been described to detect changes in protein synthesis by specific analysis of actively transcribed mRNAs (Ingolia *et al.*, 2009). However, as the method strictly speaking provides information about the association of mRNAs with ribosomes, it cannot directly quantify protein production. Regulation at the level of translation is thus typically investigated on a gene by gene basis with radioactive amino acids (e.g. ^{35}S-methionione; Pollard, 1996) or by expression of artificial reporter constructs, e.g. luciferase reporter assays (Dyer *et al.,* 2000). As this is impractical for system-wide investigations new techniques for the global analysis of cellular protein translation are required.

In recent years, mass spectrometry-based proteomics has developed to a point where it is routinely applied to tackle various pressing questions that are not or not sufficiently approachable with conventional techniques. Mass spectrometry-based proteomics can identify hundreds of proteins in complex mixtures and recently developed methodologies offer the opportunity to obtain quantitative proteomic information. Although such approaches can be harnessed to monitor global protein expression changes upon perturbations, measuring changes in protein steady-state levels upon miRNA over-expression or depletion is not ideally suited for assessing the effect on protein synthesis. Due to individual protein turnover times, a given miRNA that for instance completely shuts off protein production will rapidly alter steady-state levels of high turnover proteins while stable proteins will be affected later. In fundamental biological processes such as differentiation, however, the expression of miRNAs is strongly induced (or switched off) in a relatively small time window (Stefani & Slack, 2008). Thus, to assess endogenous regulation of mRNA translation by miRNAs a technique is needed to directly measure genome-wide changes in protein synthesis shortly after changes in miRNA expression and irrespective of individual protein turnover.

I.2.3. The complex relationship between mRNAs and proteins

The central dogma of molecular biology states that genetic information is transferred from DNA via mRNA to proteins. At each step gene expression is tightly regulated by the complex interplay of transcriptional, post-transcriptional and post-translational control mechanisms (Ben-Tabou de-Leon & Davidson, 2009). Transcription depends on *cis*- and *trans*-regulatory elements, histone modifications and chromatin structure (Kim *et al.*, 2009; Kouzarides, 2007). RNA-binding proteins and microRNAs regulate mRNA stability and translation into proteins (Ambros, 2008; Gebauer & Hentze, 2004; Keene, 2007; Sonenberg & Hinnebusch,

2009). Protein degradation is regulated by specific post-translational modifications such as phosphorylation and ubiquitination (Elsasser & Finley, 2005; Hershko & Ciechanover, 1998; Jesenberger & Jentsch, 2002; King *et al.*, 1996; Kirkpatrick *et al.*, 2005; Mayer, 2000). Although these regulatory processes are well-characterized mechanistically, little is known about how their combined effect shapes cellular gene expression.

Recently, high-throughput approaches have been used to simultaneously measure protein and mRNA concentrations (reviewed in: de Sousa Abreu *et al.,* 2009; Maier *et al.,* 2009). Most reports comparing mRNA and protein levels arrived at the conclusion that the overall correlation is poor and varies widely across organisms, with a squared Pearson's correlation coefficient (R^2) between 0.30 and 0.50. Since R^2 reflects to what extent the variation in one variable can be explained by changes in another variable, less than 50% of the variance in protein levels can be explained by differences in mRNA levels. Thus, differences in mRNA expression determined by mRNA transcription and decay are obviously insufficient to fully explain differences in protein abundance given by translation and degradation. However, the available data suffer from several limitations rendering data interpretation difficult.

First, most studies are limited to a few hundred genes, mainly due to the technical challenges involved in large scale protein quantification (Gygi *et al.,* 1999; Nie *et al.,* 2006; Tian *et al.,* 2004). Also, several studies report only relative changes in protein and mRNA concentrations and cannot be used to compare protein and mRNA levels directly (Lin *et al.,* 2005; Unwin *et al.,* 2006; Orntoft *et al.*, 2002; Berger *et al.*, 2004; Beak *et al.*, 2008; Selbach *et al.,* 2008). Absolute concentrations are certainly more difficult to obtain and hence less available than relative concentrations (cf. I.1.2.4). While genome-scale studies employing next-generation sequencing render absolute mRNA quantifications feasible, though with intrinsic limitations (Mortazavi *et al.,* 2008; Nagalakshmi *et al.*, 2010), data of comparable magnitude and quality remain elusive for proteins (Steen & Pandey, 2002). Absolute protein concentrations in yeast have been approached using libraries of TAP/GFP-fusion proteins, however, such laborious studies are confined to only very few organism (Ghaemmaghami *et al.,* 2003; Howson *et al.*, 2005; Cohen *et al.,* 2008). Alternatively, mass specrometry-based quantification of cellular protein concentrations of the human pathogen *Leptospira interrogans* has recently been described (Malmstroem *et al.,* 2009) employing synthetic stable isotope-labeled peptides as internal references. The generation of such synthetic peptides readily amenable to mass spectrometry analysis, however, requires several rounds of optimization which significantly hampers the straightforward application in other organisms (Mallick, 2007). As the ability to precisely and comprehensively measure mRNA and protein concentrations significantly impinges on the observed protein and mRNA correlation, further instrumental improvements will certainly foster the generation of data of high(er) biological relevance. For instance, in a recently developed approach termed ribosome footprinting, next-generation sequencing was applied to measure ribosome-bound, actively translated mRNAs in yeast (Ingolia *et al.*, 2009). By comparison to publicly available protein measurements, the analysis yielded a genome-wide correlation between mRNA and protein levels of $R^2 = 0.42$ which fits well to previous studies

(see above). Although the study was well-controlled, of impressive magnitude and hence of reasonable validity, it struggles with another frequently observed constraint in integrative analyses. Levels measured in one experiment are oftentimes compared to concentrations determined in a different experiment performed at a different time in a different lab, making it difficult to interpret the degree of correlation. In the high-resolution study mentioned above measurements of mRNAs obtained by ribosome footprinting were compared to measurements of protein concentrations performed several years ago using a collection of TAP/GFP-tagged yeast strains (Ghaemmaghami *et al.,* 2003). Since fluorescence- and Western blot-based detection of tagged proteins is intrinsically error-prone, the true correlation can expected to be higher. Along these lines, the correct assessment of technical and experimental error is crucial for a reasonable interpretation of integrative studies. Ideally, studies aiming for comparison of mRNA and protein levels should be designed in a way that both quantities are acquired simultaneously in the exact same experiment.

Second, several studies were performed in bacteria and neglect regulatory mechanisms specific for higher eukaryotes. In contrast to prokaryotes, transcription and translation is temporally and spatially uncoupled, i.e. eukaryotic translation takes place in the cytoplasm after transcription inside the nucleus, leading to much more elaborate regulation of eukaryotic gene expression. Upon mRNA synthesis several factors potentially impact on various steps of translation in eukaryotes. For instance, ribosomes can encounter upstream open reading frames (uORFs) situated in the mRNA's 5´UTR and change levels of translation of the main open reading frame in a competitive manner (Calvo *et al.*, 2009). Secondary structures affect translation by slowing down ribosome passage as well as a sub-optimal Kozak sequences impair translation initiation (Kozak, 1987). Also, translation is influenced by *cis*-regulatory sequence features primarily situated in 3´ untranslated regions (3´UTR) recognized by specific *trans*-acting factors, e.g. RNA-binding proteins (RBPs) and microRNAs.

Third and perhaps most importantly, conventional mRNA and protein expression profiling experiments measure steady-state concentrations which always reflect the net outcome of antagonizing synthesis and degradation processes. Since steady-state levels, however, cannot be used to disentangle the relative contributions of mRNA and protein synthesis and degradation, the actual dynamic parameters of gene expression that are subject to extensive transcriptional and post-transcriptional regulation remain elusive. To illustrate, an increase in the level of expression of a given transcript or protein can be achieved by enhanced rates of synthesis or reduced rates of degradation. To this end, mRNA and protein turnover have globally been investigated with drugs inhibiting transcription or translation, respectively, followed by quantification of mRNA and protein decay (Belle *et al.*, 2006; Raghavan *et al.,* 2002; Yang *et al.,* 2003). Classically, transcriptional inhibition by actinomycin and translational inhibition with cycloheximide has been employed. However, due to cellular toxicity of the applied drugs, turnover is measured when overall mRNA or protein synthesis is abrogated and thus may not reflect the actual turnover rate under physiological growth conditions. The stability of enzymes governing degradation itself might also be affected what further complicates the

accurate measurement of turnover rates.

In a recent study, protein stability profiling of roughly 8,000 human proteins using fluorescently tagged proteins has been described (Yen *et al.,* 2008). Intrinsic limitations of the approach arise from the fact that tagging potentially impairs protein stability by affecting folding or obscuring degradation signals. Fusion proteins may also exhibit different cellular localizations compared to their wild-type counterparts what likely impinges on protein degradation. Hence, traditional radioactive pulse-chase experiments are the gold standard method to determine half-lives of individual mRNAs and proteins (Kenney, 1967; Puckett *et al.*, 1975), albeit they are impractical for large scale studies. Recently, variants of these approaches utilizing non-radioactive tracers have been developed that allow for global analysis of mRNA and protein turnover (Pratt *et al.,* 2002; Beynon & Pratt, 2005; Gouw *et al*., 2010; Doelken *et al.,* 2008). Concurrent with recent advances in quantitative mass spectrometry, SILAC-based approaches have been developed to determine individual protein turnover. When non-labeled (light) cells are transferred to heavy SILAC growth medium, newly synthesized proteins incorporate the heavy label while pre-existing proteins remain in the light form. Consequently, the ratio of heavy (H) to light (L) peptides indicates the turnover rate of the respective protein. However, due to technical constraints given by limited sensitivity and dynamic range, protein turnover times have so far only be determined for a few hundred proteins (Doherty *et al.,* 2009).

Conceptually similar to the labeling of newly-produced proteins with stable-isotope labeled amino acids, newly synthesized RNA can be marked with the nucleoside analogue 4-thiouridine (4sU) in the mode of transcription. 4sU containing mRNA can be specifically biotinylated and recovered from total RNA by column-based affinity purification. Subsequent comparison of newly synthesized and pre-existing mRNA fractions enables global quantification of mRNA half-lives (Doelken *et al.*, 2008).

Once such quantitative measurements are available, mathematical models can be used to simulate biological systems and the underlying organizational principles (Komurov & White, 2007). Mathematical models attempting to quantitatively describe gene expression dynamics have been used for decades, though with varying success since the quality of the model and predictions thereof vastly depend on the quality of biological data used as input. However, even ordinary models built on confined experimental measurements have already proven useful in the deduction of fundamental kinetic parameters of gene expression (reviewed in: de Sousa Abreu *et al.,* 2009; Ben-Tabou de-Leon & Davidson, 2009). Specifically, with measurements of mRNA and protein levels and half-lives the prediction of transcription and translation rates becomes possible and a comprehensive description of gene expression comes within reach. Although it is clear that the combined effect of transcription, translation, mRNA decay and protein degradation shapes gene expression, their relative contributions to steady-state protein concentrations remain unknown (Vogel *et al.*, 2010). A dynamic description of gene expression will therefore help to illuminate the discrepancies observed in the complex relationship between transcriptome and proteome. However, even though recent technical advances have rendered dynamic properties amenable, a systematic comparison of those

parameters acquired from a single source has not yet been performed in any organism.

I.3. Outline and objectives of the thesis

Motivated by the lack of methods to explicitly investigate differences in protein translation across different cellular conditions on a proteome-wide scale, a modified version of the classic SILAC approach ought to be established (cf. III.1). Unlike in "classic" SILAC, the labeling in the modified version is conducted for a defined time period only (pulsed labeling), and more importantly, by means of a second heavy isotope label. As the approach exclusively labels newly-synthesized proteins, relative changes in protein translation rates between two cellular conditions can be accurately quantified.

The pulsed SILAC approach is thus ideally suited to study post-transcriptional regulation exerted by microRNAs. Over the past decade, microRNAs have ermerged as a major class of regulatory genes that play a crucial role in diverse biological processes. miRNA mutations or misexpressions are therefore correlated with various human diseases including cancer and neurological disorders. Animal microRNAs regulate gene expression by inhibiting translation and/or inducing degradation of target messenger RNAs. While microarray analyses have proven useful to assess global regulation exerted by microRNAs on the transcriptome, the impact onto the proteome given by control of mRNA stability *and* protein translation remains elusive. To this end, the pSILAC approach is applied to examine the full regulatory potential of microRNAs on a systems perspective (cf. III.2). In that context, robust microRNA target identification bears the potential to pinpoint disease-related genes which might prove useful in the diagnosis and treatment of human diseases.

Pulse labeling of newly-synthesized proteins with stable isotopes can also be utilized to analyze fundamental dynamic features of gene expression. Steady-state mRNA and protein levels are the result of a tightly controlled balance between the relative rates of production and degradation. While genome-wide profiling of both the transcriptome and proteome has been employed to elucidate relative changes in amounts upon external perturbations, it can obviously not disentangle the nature of those changes. Hence, knowledge about the relative contribution of synthesis and decay rates (commonly referred to as turnover) to gene expression is key to understand the underlying dynamics properties of the system. Though methods to study mRNA and protein turnover times are now available, a systematic comparison of mRNA and protein half-lives and abundances has not yet been performed in any organism. In order to simultaneously measure protein and mRNA turnover in mammalian cells, metabolic pulse labeling with heavy stable isotope encoded amino acids (SILAC) and the nucleoside analogue 4-thiouridine is employed (cf. III.3). Sample analysis is performed by mass spectrometry and next generation sequencing which also allows the deduction of absolute mRNA and protein levels, two parameters indispensable to quantitatively describe basic gene expression. A global characterization of mRNA and protein turnover will likely shed more light on how their combined effect shapes cellular gene expression.

II Material and Methods

II.1. General suppliers

If not stated otherwise all chemicals were purchased from Roth (Karlsruhe), Sigma-Aldrich (Steinheim) or Merck (Darmstadt).

II.2. General solutions and buffers

10x TBS

200 mM Tris base, 1.4 M NaCl, adjusted to pH 7.4

1x TBS-T

0.1 % [v/v] Tween 20 in 1x TBS

10x PCR buffer

100 mM Tris base, 500 mM KCl, adjusted to pH 8.3, stored at -20°C.

TBE

890 mM Tris base, 890 mM boric acid, 20 mM EDTA-NA2, adjusted to pH 7.4

Transfer buffer

25 mM Tris-HCl pH 8.5, 190 mM glycine, 10% methanol [v/v], 0.1% SDS [w/v]

Stripping buffer

68 mM Tris-HCl, pH 7.5, 2% SDS [w/v], 0.8% β-mercaptoethanol [v/v]

LB medium

10 g Bacto tryptone (BD), 5 g Bacto yeast extract (BD), 5 g NaCl, ad 1 L with ddH_2O

adjusted to pH 7.4; for plates 1.5% [w/v] agar was added

5x DNA loading buffer

50 % glycerol (85%) [v/v], 0.5 x TBE, 0.2% SDS [w/v], dash bromphenol blue

Radioimmunoprecipitation assay (RIPA) buffer

50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1mM EDTA, 1% [v/v] Triton-X 100, 1% [w/v] Na-Deoxycholate, 0.1% [w/v] SDS. Complete Protease Inhibitor cocktail (Roche) (1:25) and Phosphatase Inhibitor Mix 1 (Sigma) (1:100) were added.

Fixative solution

50% methanol [v/v], 10% acetic acid [w/v]

Blocking solution

1x TBS-T supplemented with 5% [w/v] non-fat dry milk powder

II.3. Solutions and buffers for LC-MS/MS sample preparation and instrumentation

ABC buffer

50 mM ammonium bi-carbonate (NH_4HCO_3, ABC) in water (pH 8.0). Stored at RT.

Destaining buffer

25 mM NH_4HCO_3 / 50% ethanol (EtOH). Equal volumes of digestion buffer and 100% EtOH were combined and stored at RT.

Reduction buffer

10 mM dithiothreitol (DTT) in 50 mM ABC buffer. Stored in small aliquots at -20°C.

Alkylation buffer

55 mM iodoacetamide in 50 mM ABC buffer. Stored in small aliquots at -20°C and kept in the dark.

Extraction buffer

3% trifluoroacetic acid (TFA) / 30% acetonitrile (ACN). Stored at RT.

Denaturation buffer

6 M urea/2 M thiourea in 10 mM HEPES (pH 8.0). Stored in small aliquots at – 80°C.

Buffer A

5% acetonitrile, 0.1% formic acid in water (LiChrosolv grade, Fischer Scientific). Stored at RT.

Buffer B

0.1% formic acid, 80% acetonitrile in water (LiChrosolv grade, Fischer Scientific). Stored at RT.

II.4. Cell culture and preparation of SILAC medium

II.4.1. Cell culture for pulsed SILAC (pSILAC)

For iron treatment experiments, HeLa cells obtained from LGC Promochem were cultivated in SILAC medium at 37°C with 5% CO_2 and split every second or third day.

For luciferase induction experiments, HeLa Tet-On cells with an inducible luciferase transgene were used. HeLa M2 cells expressing a modified version of a reverse tetracycline controlled transactivator have been described previously (Hampf & Gossen, 2007). These cells were stably transfected with a firefly luciferase gene under control of pTET, a promoter responsive to tetracycline controlled transactivators (Gossen & Bujard, 1992). The resulting clonal double stable cell line therefore shows a Tet-On type response, as luciferase expression is induced upon addition of the tetracycline derivate doxycycline.

Yeast haploid strain BY4742 (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) strain was grown in SILAC YPD liquid medium at 30°C and 220 rpm.

II.4.2. Cell culture for determination of mRNA and protein half-life

For half-life determination experiments NIH3T3 (mouse fibroblasts, obtained from ATCC) cells were used and cultivated in SILAC medium at 37°C with 5% CO_2 and split every second or third day.

II.4.3. Cell culture for microRNA/LNA transfection experiments

All experiments were performed with HeLa cells obtained from LGC Promochem. Cells were cultivated in SILAC medium at 37°C with 5% CO_2 and split every second or third day.

II.4.4. Cell culture for intensity based absolute quantification (iBAQ)

E.coli strain DH10B (*F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu) 7697 galU galK rpsL (StrR) endA1 nupG λ-*) was grown over night in Luria-Bertani (LB) broth at 37°C and 250 rpm.

NIH3T3 cells were cultivated at 37°C with 5% CO_2 in Dulbecco's Modified Eagle's Medium (DMEM) Glutamax (Gibco) supplemented with 10% fetal bovine serum (Gibco) and harvested at 70% confluence. NIH3T3 cell numbers were determined using haemocytometers (C-Chips™) and protein concentration was measured using Bradford assay (Pierce).

II.5. Preparation of SILAC media

Preparation of SILAC media was essentially done as described previously (Ong &Mann, 2006)For preparation of SILAC media for mammalian cells, Dulbecco's Modified Eagle's Medium (DMEM) Glutamax lacking arginine and lysine (a custom preparation from Gibco) supplemented with 10% dialyzed fetal bovine serum (dFBS, Gibco) was used. "Heavy" and "medium-heavy" SILAC media were prepared by adding 84 mg/l $^{13}C_6{}^{15}N_4$ L-arginine plus 146 mg/l $^{13}C_6{}^{15}N_2$ L-lysine or 84 mg/l $^{13}C_6$-L-arginine plus 146 mg/l D_4-L-lysine, respectively. Labeled amino acids were obtained from Sigma Isotec ($^{13}C_6$-L-arginine, $^{13}C_6{}^{15}N_4$ L-arginine and $^{13}C_6{}^{15}N_2$ L-lysine) and Cambridge Isotope Laboratories (D_4-L-lysine). To prepare "light" SILAC medium, the corresponding non-labeled amino acids (Sigma) were added.

Yeast haploid strain BY4742 (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) strain was completely SILAC labeled in the traditional way by cultivating them over night in minimal media (cf. Table II.1) supplemented with light (L), medium-heavy (M) or heavy lysine (H).

Table II.1: SILAC yeast culture medium[1]

Component	Catalog No.	Solvent	Final concentration
YNB without amino acids and ammonium sulfate	Y1251	Milli-Q water	1.7 g/l
D-Glucose	G7528	Milli-Q water	20 g/l

Ammonium sulfate	A4418	Milli-Q water	5 g/l
Adenine hemisulfate	A9126	Milli-Q water	200 mg/l
L-tyrosine	T3754	0.5 M NaOH	100 mg/l
L-histidine·HCl·H2O	H8125	0.5 M NaOH	10 mg/l
L-leucine	L8000	0.5 M HCl	60 mg/l
L-methionine	M9625	PBS	10 mg/l
L-phenylalanine	P2126	0.5 M HCl	60 mg/l
L-tryptophane	T0254	0.5 M NaOH	40 mg/l
Uracil	U0750	Milli-Q water	20 mg/l
L-arginine·HCl[2,3]	A5131	PBS	100 g/l
L-lysine·HCl[4,5]	L5501	PBS	146 g/l

1 Filter-sterilize the medium through a 0.22 µm filter (Millipore Express PLUS filter membrane, Billerica, MA, USA). Store at 4°C.

2,3 To prepare light (L) medium, L-arginine-$^{12}C_6$·HCl (Arg-0) was used. For medium-heavy (M) and heavy (H) medium, L-arginine-$^{13}C_6$·HCl (Arg-6) and L-arginine-$^{13}C_6{}^{15}N_4$·HCl (Arg-10) were used, respectively.

4,5 To prepare "light" medium, L-lysine-$^{12}C_6{}^{14}N_2$·HCl (Lys-0) was used. For "medium-heavy" and "heavy" medium, D_4-lysine (Lys-4) and L-lysine-$^{13}C_6{}^{15}N_2$·HCl (Lys-8) were used, respectively. Arg-6 (Cat.# 643440), Arg-10 (Cat.# 608033), Lys-4 (Cat. #616192) and Lys-8 (Cat.# 608041) were obtained from Sigma Isotec; all the other chemical components were from Sigma-Aldrich.

II.6. Quantification of luciferase expression by pulsed SILAC and luminescence

In order to validate the pulsed SILAC approach, mass spectrometry-based quantification of gene expression was compared with the commonly used luciferase reporter system. A 10 h time course experiment was designed that allowed for parallel measurement of SILAC- and luminescence-based relative quantification of luciferase expression. Fourteen 10 cm dishes with TET-On cells grown to 80 % confluency in DMEM Glutamax were prepared. Cells from seven of these dishes were washed with D-PBS (Gibco) and transferred to medium-heavy (M) SILAC medium. Simultaneously, luciferase expression was induced by adding 500 ng/ml doxycycline to the cell culture medium followed by incubation for the indicated lengths of time (1, 2, 4, 6, 8 and 10 h). In parallel, the remaining seven cell culture dishes were washed and transferred to heavy (H) SILAC medium, treated with 500 ng/ml doxycycline and incubated for 10 h. These cells reflect maximum luciferase expression after 10 h of induction and serve as an internal reference for SILAC-based relative quantification.

Cells were lysed with 500 µl PBS containing 0.5 % Triton X-100 and complete protease inhibitors w/o EDTA (Roche) for 10 minutes on ice. The lysates were cleared for 10 min by centrifugation (14,000 rpm at 4°C). 10 µl aliquots of medium-heavy (M) sample were mixed with 100 µl of 25 mM glycylglycine, 15 mM $MgSO_4$, 5 mM ATP and 0.05 mM luciferin and assayed for luciferase activity in a microplate luminometer (TR717 Tropix Perkin Elemer, Applied Biosystems) using the integral mode (10 s). The background signal measured in extracts of TET-On HeLa cells that were not induced by Dox (0 h sample) was indistinguishable from instrumental background (100-150 relative light units (rlu)/10 s). The luminescence-based relative quantification of

luciferase expression was performed by referring the absolute luminescence intensities for the different incubation times (labeled medium-heavy (M)) to the 10 h time point (control cells, indicating the maximum gene expression and maximum luminescence intensity, heavy (H) labeled). For pulsed SILAC-based quantification, lysates from the seven dishes transferred to heavy (H) cell culture medium for 10 h were pooled and mixed in a 1:1 ratio with the medium-heavy (M) lysate of each indicated sample time point. The mixed lysates were concentrated using Vivaspin columns (Vivascience AG) with a 3 kDa cut-off, supplemented with 4x SDS-sample buffer and separated by SDS-PAGE (Novex 4-12 % gradient gel, Invitrogen). The gel was stained with Coomassie Blue and gel slices in the expected molecular weight range of the luciferase protein (62 kDa) were cut out.

II.7. Treatment of HeLa cells with iron (FAC) or iron chelator (DFO)

Two 10 cm dishes of HeLa cells were grown to 60 % confluency in DMEM Glutamax. Cells from one dish were washed with PBS and transferred to heavy (H) SILAC medium with simultaneous addition of ferric ammonium citrate (FAC) or deferoxamine (DFO) to a final concentration of 100 µM. In parallel, cells from the second dish were washed and transferred to medium-heavy (M) SILAC medium without additional iron. After 14 h of incubation, the cells from both dishes were scraped off, combined and spun down (10 min, 600 g, 4°C). Cell pellets were lysed in SDS sample buffer, separated by SDS-PAGE (Novex 4-12% gradient gel) and processed by in-gel digestion using standard protocols (12 gel slices in total).

The effect of iron has also been investigated in a time-course experiment to allow for a time-resolved acquisition of protein changes. FAC treatment and SILAC labeling were performed as described above and cells were harvested 4 h, 6 h, 8 h and 12 h upon perturbation.

II.8. microRNA or LNA transfection and pulsed SILAC labeling

HeLa cells were transfected with synthetic miRNAs (Dharmacon) or locked nucleic acids (LNAs) (BioTez) using DharmaFECT1 (Dharmacon) according to the manufacturer's instructions. Cells were plated in 15 cm dishes in light (L) SILAC medium the day before transfection to a confluency of 30-40%. Synthetic miRNAs/LNAs were used at a final concentration of 100 nM. Control transfections were carried out in parallel under the same conditions but using water instead of the miRNAs/LNAS (MOCK-transfection). At 8 h post transfection, cells were trypsinized. 2/5 of the cell suspension was used for RNA isolation (8 h time point). The remaining cells were transferred into two new 10 cm dishes each (3/10 of all cells per plate). One of the two plates containing miRNA/LNA transfected cells was transferred to medium-heavy (M) SILAC medium and one of the MOCK-transfected cells to heavy (H) SILAC medium for pulsed SILAC labeling. The two remaining plates were kept in normal light (L) medium for mRNA analysis. After 24 h, cells of the two SILAC plates were scraped off, combined and spun down (10 min, 600 x g, 4°C) for protein analysis. The corresponding plates were harvested for total RNA

isolation (32 h time point). Reproducibility was checked by performing two independent miR-1 transfection experiments on different days. Both samples were processed and analyzed by mass spectrometry (each on 15 slices measured in triplicates). 2,287 proteins were identified and quantifiable according to the applied quantification criteria after removal of 5% outliers (Grubb´s test).

II.8.1. Synthetic miRNAs

Synthetic miRNAs mimicing mature endogenous miRNAs were obtained from Dharmacon as annealed, 2'-deprotected and desalted duplexes. miR-1 was exactly designed as described in the study of Lim *et al.* 2005, that is carrying one mismatch in the duplex to facilitate activation of the sense strand (Lim *et al.,* 2005; Khvorova *et al.,* 2003).

RNA duplexes were synthesized as follows (sense 5'-3' / antisense 3'-5'):

miR-1: UGGAAUGUAAAGAAGUAUGUAA / AUAACUUACAUUUCUUCAUACA

miR-16: UAGCAGCACGUAAAUAUUGGCG / AUAUCGUCGUGCAUUUAUAACC

miR-30a: UGUAAACAUCCUCGACUGGAAG / CGACAUUUGUAGGAGCUGACCU

miR-155: UUAAUGCUAAUCGUGAUAGGGGU / ACAAUUACGAUUAGCACUAUCCC

let-7b: UGAGGUAGUAGGUUGUGUGGUU / CCACUCCAUCAUCCAACACACC

II.8.2. Locked nucleic acids (LNAs)

LNAs (locked nucleic acids) (Wahlestedt *et al.*, 2000; Stenvang & Kauppinen, 2008) purchased from BioTez were designed to bind endogenous mature miRNAs with perfect complementarity. The following LNAs (5'-3') were used in this study:

LNA-anti-let-7b: AACCACACAACCTACTACCTCA

LNA-anti-miR-21: TCAACATCAGTCTGATAAGCTA (used as control for LNA-anti-let-7b luciferase experiment)

II.9. Analysis of transfection efficiency

The BLOCK-iT fluorescent oligo (Invitrogen) is a fluorescein-labeled, non-targeted dsRNA oligomer allowing for visual monitoring of transfection efficiency. Cells were transfected with BLOCK-iT fluorescent oligos as described for miRNAs/LNAs. 8 h post transfection cells were washed with 1x PBS (Gibco) and fixed with 4% paraformaldehyde (PFA) in PBS. Transfection efficiency was assessed by comparing the amount of transfected, fluorescing cells with non-transfected cells using epifluorescence microscopy (Leica DM-R).

II.10. Mapping of protein and mRNA IDs and identification of sequence motifs correlated with changes in protein production

Microarray data analysis as described below was conducted by Zhou Fang, bioinformatician in the group of Nikolaus Rajewsky at the Berlin Institute for Medical Systems Biology (BIMSB, Robert-Rössle Str. 10, D-13125 Berlin, Germany):

II.10.1. Mapping

Protein identifiers were first mapped to NCBI Entrez Gene gene numbers by the IPI cross-reference file (version 3.37) from EBI database (http://www.ebi.ac.uk). The gene2refseq file (downloaded on 28th of Sep, 2007) from the NCBI database (ftp://ftp.ncbi.nlm.nih.gov) was used to map NCBI Entrez Gene gene numbers to Refseq identifiers. For each protein group, the first protein identifier in a group was assigned to corresponding Refseq identifiers. If the first protein did not have any corresponding Refseq identifiers the mapping was done for the second protein in the protein group.

II.10.2. Identification of correlated sequence motifs

Human 3' UTR, 5' UTR and CDS sequences based on human reference sequence (NCBI Build 36.1) were extracted from UCSC Genome Browser (http://genome.ucsc.edu). In order to assign characteristics (number of seeds) to unique protein identifiers, the maximum number of seeds from all transcript identifiers mapped to the protein was used. Then, a linear regression model was employed to identify significant motifs in mRNA 3' UTRs, 5´UTRs and coding sequences which correlated best with the global changes in protein synthesis (Bussemaker *et al.,* 2001; Sood *et al.,* 2006).

II.11. Generation of luciferase reporter constructs

If not stated otherwise, standard methods were used for agarose gel electrophoresis, isolation of plasmids by alkaline lysis, polymerase chain reaction (PCR), DNA digestion by restriction enzymes and ethanol/isopropanol precipitation. The equipment used is up to latest laboratory standard.

For pSILAC validation, luciferase reporters carrying 3' UTRs of genes found to be down-regulated by pSILAC upon microRNA over-expression were constructed. The 3' UTRs were PCR-amplified from HeLa cDNA (purchased from BioCat, Catalog No.: C1255811) and cloned into *Xho*I (NEB) and *Not*I (NEB) sites immediately downstream of the stop codon in the pRL-TK CXCR4 4x vector (a kind gift of Phil Sharp) coding for Rr-luc. The artificial CXCR4 4x target site has been removed by digestion beforehand. All constructs were checked by sequencing.

II.11.1. 3' UTRs of the following genes were selected for cloning

c-Met (NM_000245), RDH10 (NM_172037), CAP1 (NM_006367), TAGLN2 (NM_003564), ADPGK (NM_031284), MTX1 (NM_002455), SLC25A1 (NM_005984), ATP6V0A1 (NM_005177)

II.11.2. Primers (5'-3')

All synthesized primers (BioTez) contain flanking restriction sites (5'-3' FW primer = *Xho*I site; 5'-3' RW primer = *Not*I site) for site-directed insertion of the PCR product into the target vector.

Table II.2: Primers for generation of reporter constructs

Gene	Accession ID	Sequence 5´->3´
CMET	NM_000245_FW	CGGCTCGAGTGCTAGTACTATGTCAAAGCAA
CMET	NM_000245_RV	ATAGTTTAGCGGCCGCTGCATGATTTATCAGAACAACT
RDH10	NM_172037_FW	CGGCTCGAGGAATCTTTTTGTATGGAATATT
RDH10	NM_172037_RV	ATAGTTTAGCGGCCGCCAGTCATTTATAAAACTCCCCA
TAGLN2	NM_003564_FW	CGGCTCGAGTCCCACCCCAGGCCTTGCCC
TAGLN2	NM_003564_RV	ATAGTTTAGCGGCCGCCAAAAATGACAAATTCTTTA
MTX1	NM_002455_FW	CGGCTCGAGTTTGTCCTCACGCTCCCAAG
MTX1	NM_002455_RV	ATAGTTTAGCGGCCGCCAGTGTGAGTGGCTTTATTC
SLC25A1	NM_005984_FW	GCTCTAGAGCCTAGAGAGGCCGCAAGGG
SLC25A1	NM_005984_RV	ATAGTTTAGCGGCCGCGCAACAGGATCCGGTTTATT
CAP1	NM_006367_FW	CGGCTCGAGGCGAAGTGCCACTGGGTTCT
CAP1	NM_006367_RV	ATAGTTTAGCGGCCGCCAAGTTTGGTATTAACTTTA
ADPGK	NM_031284_FW	CGGCTCGAGGAAGATTCTTAGGGGTAATT
ADPGK	NM_031284_RV	ATAGTTTAGCGGCCGCCCTGAAATGTAAATTGTTTT
ATP6V0A1	NM_005177_FW	CGGCTCGAGGTCCCTGTGAGGGCCGTGTG
ATP6V0A1	NM_005177_RV	ATAGTTTAGCGGCCGCCCGGGGAAGTCAAACATACT

II.11.3. Polymerase chain reaction (PCR)

Standard polymerase chain reaction was used to amplify 3´UTR cDNA fragments for cloning. PCR reaction master mix was assembled on ice (50 ng Template; 50 pmol forward primer; 50 pmol reverse primer; 5 µl 10x Taq-buffer (NEB); 1 µl dNTP Mix 10 mM; 2.5 U Taq DNA polymerase (NEB); ad. 50 µl ddH_2O). PCR was performed using a Gene Amp 2400 cycler (Perkin-Elmer) according to the following program: first denaturation 2 min 94°C, followed by 35 cycles of 30 s denaturation at 94°C, 60 s of annealing at 55°C, 50 s of elongation at 72°C and a final polymerization step of 7 min at 72°C.

II.11.4. Agarose gel electrophoresis

For the preparative analysis of DNA fragments obtained by PCR or excision of specific DNA fragments after restriction reactions, gel electrophoresis was performed. Separation is achieved as DNA fragments migrate according to their size differently along an electrical field in the

agarose gel. Separation and detection of DNA fragments in the gel was performed with 0.4-1.2 % [w/v] of agarose in 0.5x TBE buffer supplemented with 1 µg/mL ethidiumbromide and applying 80-140 V for the aspired time. Prior to loading the samples on the gel, probes were mixed with DNA loading buffer. Visualization of DNA was achieved using a Gene Genius device (Syngene) thereby detecting the fluorescence of the DNA-intercalating dye ethidiumbromide. To identify size of nucleic acids, standard markers (DNA ladder 1kb or 50 bp; Fermentas) were used. DNA fragments of interest were eluted from agarose gels using the QIAquick gel extraction kit (Qiagen) according to manufacturer's instructions.

II.11.5. Enzymatic DNA digestion

In order to open or cut out DNA fragments from the plasmid pRL-TK CXCR4 4x and also to prepare PCR products for ligation reactions the restriction enzymes *Xho*I and *Not*I (all from NEB) were used. DNA was purified before digestion using a PCR purification kit (Qiagen). Enzymes were diluted in appropriate restriction buffers and digestion was allowed to proceed at 37°C for 2 h. Reaction products were purified again by gel electrophoresis.

II.11.6. Ligation of DNA fragments

Ligations of DNA fragments with the target vector were carried out by employing the Rapid DNA Ligation Kit (Roche). For each mixture, 10 µL of pre-aliquoted ligation buffer containing ATP was used. A molecular fragment:plasmid ratio of 3:1 was employed and the ligation reaction was incubated for 15-30 min at RT, transferred on ice and then purified using the PCR purification kit (Qiagen).

II.11.7. Preparation of plasmid DNA

For isolation of plasmid DNA from bacteria a mini or midi prep kit from Qiagen was used according to the manufacturer's instructions. The DNA concentration was measured spectrometrically using a Nanodrop ND-1000 device (Peqlab).

II.11.8. DNA Sequencing

DNA sequencing was performed on-campus by Invitek DNA sequencing service (Berlin-Buch).

II.12. Co-transfection of synthetic miRNAs and 3' UTR reporter constructs

HeLa cells were co-transfected with synthetic miRNAs and different 3' UTR luciferase reporter constructs using Lipofectamine 2000 (Invitrogen) following the manufacturer's instructions (Lipofectamine™ 2000 siRNA - plasmid co-transfection protocol). The day before transfection, cells were seeded in 24-well plates in light (L) SILAC medium (1x 10^5 cells/well). The following day the 80-90% confluent cells were transfected with 180 ng of the respective reporter

plasmid and 20 ng pGL3 control plasmid (Promega). Synthetic miRNAs were co-transfected at a final concentration of 100 nM. DNA, RNA and Lipofectamine 2000 were diluted in serum-free DMEM upfront. All transfections were performed in triplicates. Control transfections were performed with miR-155 as a control since this miRNA did not significantly affect synthesis of the tested proteins (data not shown). On the next day the medium was changed and cells were harvested 48 h post transfection in 100 µl 1x Passive Lysis Buffer (Promega) according to the manufacturer's instructions. Cell lysates were cleared by centrifugation in a microcentrifuge for 5 min at 16,000x g, 4°C.

II.13. Dual-Luciferase Assay

Dual-Luciferase Assays (Promega) were performed 48 h post transfection following the manufacturer's protocol and detected with a MicroLumat Plus LB 96V luminometer (Berthold Technologies). Each cell lysate was measured three times (three technical replicates) in a 96-well plate (Nunc). Renilla luciferase activity of the pRL-TK reporter constructs was normalized to the activity of the firefly luciferase of the pGL3 control plasmid (Promega) which served as internal transfection control. Errors bars were calculated as follows: First, relative errors of the three biological replicates of the respective reporter and its corresponding control were computed. Second, the relative error of the reporter and the control were added up according to the law of error propagation. The resulting relative error was used to calculate absolute errors of the normalized expression values. To estimate the pSILAC error we calculated the standard deviation of all protein quantification for two biological replicates of the miR-1 transfection experiment after 5% outlier removal (Grubb´s test). Error bars are shown as +/- two standard deviations.

II.14. Immunoblotting (Western blotting)

II.14.1. Validation of putative miRNA targets

In order to validate selected miRNA targets identified by the pulsed SILAC approach, HeLa cells were transfected with either one of the synthetic miRNAs or the LNA targeting let-7b as described above. 32 h post transfection cells were harvested. For microRNA target validation, primary antibodies against CEBPβ (sc-150), Kras (sc-30), Annexin2 (sc-48397), Tropomyosin (sc-28543), Twinfilin-1 (sc-51241), FGF-2 (sc-74412), Integrin α2 (sc-53353), α-Adducin (sc-25731), EGFR (sc-03), PICALM (sc-6433), SNX6 (sc-50373), IMP-1(sc-21026) and cMet (sc-161) were purchased from Santa Cruz Biotechnologies or Sigma-Aldrich (β-Actin, A5441).

II.14.2. Validation of protein half-lives

For protein half-lives validation, cellular protein degradation was monitored after blockage of *de novo* protein synthesis via cycloheximide treatment. Primary antibodies against CEBPβ (sc-150),

JNK (sc-7345), Hells (sc-46665), JunB (sc-46) were purchased from Santa Cruz Biotechnologies or Cell Signaling (GSK3β, 9315) or Sigma-Aldrich (β-Actin, A5441) or BD Biosciences (556470, CCND1).

II.14.3. Detection of proteins

Harvested cells were lysed in appropriate amounts of RIPA buffer for 20 min on ice. The lysates were cleared by centrifugation for 10 min (14,000 rpm at 4°C) and transferred to fresh tubes. SDS-PAGE was performed with whole-cell lysates using NuPAGE Novex 4 to 12% gradient gels (Invitrogen) under reducing conditions according to the manufacturer's instructions. Gels were blotted onto PVDF (Perkin-Elmer) membranes either using the iBlot dry blotting system (Invitrogen) according to the manufacturer's instructions or the XCell II Blot Module (Invitrogen). In the latter case, 1 mA of current per cm2 of membrane for 2 h was applied and the membrane was activated prior to use with 100% methanol and washed once in transfer buffer. Unspecific binding sites were blocked with 5% dry milk powder in 1x T-BST at RT for 30 min followed by washing three times for 10 min with 1x T-BST. In order to visualize specific protein bands, the membrane was incubated with the respective primary antibody at a 1:1,000 dilution in blocking solution at 4°C overnight with shaking. Blots were washed 3x in T-BST and incubated either with an anti-mouse, an anti-rabbit or an anti-goat secondary antibody (all from Amersham) conjugated to horseradish peroxidase diluted 1:3,000 in T-BST for 1 h at RT. After three more washing steps in TBST for 10 min each the signal of the bound secondary antibodies were detected by applying Western Blot Chemiluminescence Reagent Plus for ECL immunostaining (PerkinElmer) to the membrane followed by exposition to X-ray films (GE Healthcare). For sequential detection of different proteins on the same membrane, the PVDF membrane was stripped in stripping buffer (50 °C, 20 min), washed once with isopropanol and thoroughly with distilled H_2O. Subsequently, the membrane was blocked again and treated as described above.

II.15. Microarray data analysis

Microarray data analysis as described below was conducted by Zhou Fang, bioinformatician in the group of Nikolaus Rajewsky at the Berlin Institute for Medical Systems Biology (BIMSB, Robert-Rössle Str. 10, D-13125 Berlin, Germany):

The output of microarrays was normalized by the standard rma() function from the Bioconductor R-library (www.bioconductor.org). To annotate Affymetrix probe sets to Refseq identifiers, the current NetAffx Annotation file was downloaded from the Affymetrix website (http://www.affymetrix.com). For the transcript (RefSeq) with multiple probes, the average logarithm expression values for all corresponding probes were taken. Fold-changes were defined as differences between the intensities of misexpressions and controls (log2 ratios). The MOCK-transfected control corresponding to the same miRNA transfection experiment was used where applicable (three control samples for miR-1 and two samples for miR-30a). For the

other miRNA transfection experiments the median of three controls or two controls taken at the 8 h and 32 h time point were used, respectively.

II.16. RNA isolation

Total RNA was extracted using Trizol Reagent (Invitrogen) following the manufacturer's instructions. Cells were washed once with 1x D-PBS (Gibco) and cells were harvested by trypsinization. Alternatively, the culture flask/dish was put on ice, cells were scraped off using a cell scraper and resuspended in ice-cold D-PBS. Cells were pelleted by centrifugation for 5 min at 300x g, 4°C with a Heraeus Multifuge 3 S (Heraeus). The supernatant was carefully removed and the cells were lysed by adding 1 ml Trizol Reagent. To ensure homogenization, the cell lysate was passed 8-10 times through a 20G needle. After adding 200 ul of chloroform (Invitrogen), the mixture was transferred to a Phase-Lock-Gel tube (5Prime), vigorously mixed and centrifuged at full speed (20,000 g) for 5 min at 4°C. The upper, RNA containing aqueous phase was removed and transferred to a fresh, RNase-free tube (Ambion). To precipitate RNA, 1/10 reaction volume of 5 M NaCl and an equal volume of 2-propanol were added and incubated for 10 min at RT. Precipitated RNA was collected through centrifugation at full speed (20,000 g) for 30 min at 4°C. The pellet was washed with an equal volume of 75% ethanol and precipitated again at 20,000g for 20 min. RNA pellets were resuspended in 20 µl RNase-free sterile water and RNA quantity was assessed spectrophotometrically using the NanoDrop ND-1000 UV-VIS Spectrophotometer (Thermo Fisher).

II.17. Double pulse-labeling of NIH3T3 cells with heavy amino acids and 4-thiouridine

After growing NIH3T3 cells to 60-70 % confluency in light (L) SILAC medium, the fibroblasts were thoroughly washed three times with pre-warmed PBS and transferred to heavy (H) SILAC medium. Only SILAC medium of cells designated for RNA isolation was additionally supplemented with 4-thiouridine (4sU) to a final concentration of 400 µM. Pulse-labeling of proteins was performed for 1.5, 4.5 and 13.5 h, while labeling of newly transcribed RNA was conducted for 2 h. At the indicated time points, cells were scraped off, spun down (10 min, 600 g, 4 °C) and used either for protein or RNA isolation.

II.18. Preparation of newly-synthesized and pre-existing RNA

Apart from minor changes, all steps described in the following were essentially performed as described previously (Doelken *et al.,* 2008):

II.18.1. Biotinylation and purification of 4sU-labeled RNA

Biotinylation of 4sU-labeled RNA was performed using EZ-Link Biotin-HPDP (Pierce) dissolved in dimethylformamide (DMF) at a concentration of 1 mg/mL and stored at 4°C. Biotinylation

was carried out in RNase-free water (Invitrogen)) with 10 mM Tris-HCl (pH 7.4), 1 mM EDTA and 1 mg/ml Biotin-HPDP (Perbio) at a final RNA concentration of 1 µg/µl RNA for 3 h in the dark at RT. In general 500 µg total RNA were used for the biotinylation reaction.

To purify biotinalyted RNA from excess of Biotin-HPDP, a Phenol:Chloroform:Isoamylalkohol (v/v = 25:24:1, Invitrogen) extraction was performed. Phenol:Chloroform:Isoamylalkohol was added to the reaction mixture in a 1:1 ratio, transferred to a Phase-Lock-Gel tube (5Prime), vigorously mixed and centrifuged at full speed (20,000 g) for 5 min at 4°C. The upper, RNA containing aqueous phase was removed and transferred to a fresh, RNase-free tube (Ambion). To precipitate RNA, 1/10 reaction volume of 5 M NaCl and an equal volume of 2-propanol were added and incubated for 10 min at RT. Precipitated RNA was collected through centrifugation at full speed (20,000x g) for 30 min at 4°C. The pellet was washed with an equal volume of 75% ethanol and precipitated again at 20,000x g for 20 min. Finally, RNA was reconstituted in 25-50 ul of RNase-free water.

II.18.2. Separation of biotinylated 4sU-RNA from unlabeled RNA

For specific isolation of biotinylated 4sU-RNA streptavidin coated magnetic beads (µMACS, Miltenyi) were used. To avoid unfavorable secondary RNA structures which potentially impair the binding to the beads, the RNA was first denatured at 65°C for 10 min followed by rapid cooling on ice for 5 min. 4sU-biotinylated RNA was incubated with µMACS streptavidin beads (usually 20 µg labeled RNA with 200 µl beads) for 15 min at RT with rotation. Actual separation of 4sU-RNA from unlabeled RNA was achieved by applying the mixture onto an µColumn (Miltenyi) in the magnetic field of the µMACS separator (Miltenyi). Before, µColumns were equilibrated with 100 µl equilibration buffer that is part of the Miltenyi µBeads Kit. To recover unlabeled, pre-existing RNA the column flow-through (FT) has been collected. Then, µColumns were washed three times with 0.5 mL 55°C washing buffer (100 mM Tris-HCl, pH 7.4, 10 mM EDTA, 1 M NaCl, 0.1% Tween20). All washing fractions were discarded. Elution of labeled RNA from the beads was achieved by adding 100 µl of freshly prepared 100 mM dithiothreitol (DTT) onto the column followed by a second elution round 3 min later with the columns being retained in the magnetic field. Lastly, RNA from the FT and elution step was recovered using the RNeasy MinElute kit and reconstituted in 25-50 µl of RNase-free water.

II.18.3. Determination of separation efficiency of 4sU-labeled RNA

For combined 4sU and radioactive RNA labeling the medium was supplemented with 125 nM 5-3-H-cytidine (25 Ci/mmol, Sigma) and 400 µM 4sU. In order to remove unincorporated 5-3-Cytidine, RNA was extracted using the RNeasy kit (Qiagen). Column-based separation was performed as described above. Radioactivity of the column flow-through, the washing fractions (all three washing fractions were pooled), the elution fractions (the two elution fractions were pooled) and of the beads was determined with a Tri-Carb Liquid Scintillation Counter (PerkinElmer). Radioactivity measured for the different fractions was summed up and set to 100% (input).

II.18.4. Estimation of the doubling time of NIH3T3 cells

To estimate the cell cycle time of NIH3T3 cells during the double pulse-labeling of RNAs and proteins, two 15 cm culture dishes with cells transferred to “heavy” SILAC medium were pooled and counted (n = 6) with haemocytometers (C-Chips™) at 0.75 h, 5 h, 9 h, 13.25 h after the onset of the experiment. The doubling time was estimated assuming exponential growth given by the following equation:

$N(t) = N_0 2^{t/tcc}$

N(t) is the number of fibroblasts at time t, N_0 the number of cells at the onset of the experiment and t_{cc} denotes the doubling time. Final cell cycle duration was calculated by applying linear regression to individual cell cycle times obtained at the respective time points.

II.19. mRNA sample preparation and sequencing

mRNA sample preparation for sequencing was performed by Na Li, PhD student in the group of Wei Chen at the Berlin Institute for Medical Systems Biology (BIMSB, Robert-Rössle Str. 10, D-13125 Berlin, Germany):

Total RNA from bulk RNA, pre-existing RNA and newly-synthesized RNA were subjected to Solexa deep-sequencing, following the mRNA sequencing protocol provided by Solexa. Briefly, poly(A) RNA was purified from 10 µg of total RNA with two rounds of hybridization to Dynal oligo(dT) beads (Invitrogen). The poly(A) RNA was first fragmented by using 5x fragmentation buffer (200mM Tris acetate, pH 8.2, 500 mM potassium acetate and 150 mM magnesium acetate) and heating at 94 °C for 3.5 min in a thermocycler. The fragmented RNA was precipitated and used for first- and second- strand cDNA synthesis with random hexamer primers. Both ends of the cDNA fragments were then repaired by T4 polymerase and Klenow DNA polymerase, and T4 polynucleotide kinase. Using Klenow exo (3´ to 5´ exo minus), a protruding “A” base was added to the 3´ end of the DNA fragments for ligation with Solexa adaptors (with a “T” overhang). After ligation, the cDNA with size of 200 +/- 25 bp was selected and followed by 15 cycles of PCR amplification. Finally the adapter-ligated DNA was sequenced for 36 cycles on the Solexa/GAIIX, according to the manufacturer’s instructions.

II.20. Calculation of mRNA half-lives

Analysis of sequencing results and calculation of mRNA half-lives were performed by Wei Chen, head of the scientific genomics platform at the Berlin Institute for Medical Systems Biology (BIMSB, Robert-Rössle Str. 10, D-13125 Berlin, Germany):

The Solexa sequencing reads were mapped to mouse genome reference sequence (NCBI Build 37/mm9, July 2007) by using SOAP2 with a maximum of 2 mismatches allowed. Only uniquely mapped reads were retained. The mRNA abundance was estimated as the number of reads mapped within the exonic regions divided by the total number of reads uniquely mapped on the genome (in Million) and the cumulated exon length (in Kb) (RPKM). Only the transcripts

with RPKM values > 1 in all three samples were retained for further analysis. Newly synthesized transcripts containing more 4sU molecules are more prone to be pulled down and thus to be subsequently sequenced. To correct for such a bias in pulldown efficiency of mRNAs containing different numbers of uridines, local regression with the LOWESS method was applied (cf. Fig. II.1):

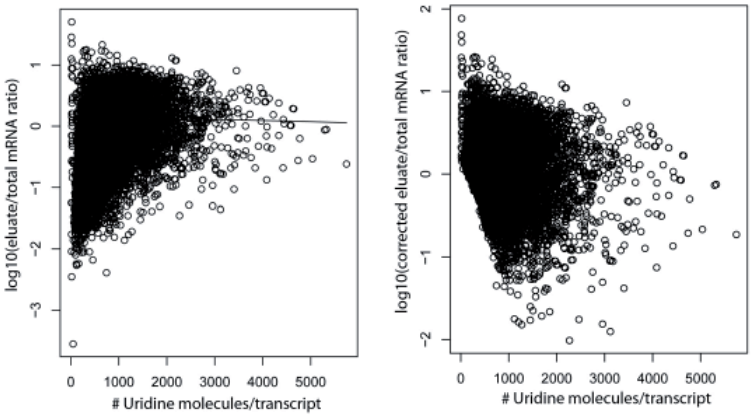


Fig. II.1 | Correction of 4sU bias by LOWESS regression. Plotting the number of uridines per transcript against newly-synthesized/ total RNA ratios obtained by RNA-Seq reveals the bias in pulldown efficiency (left plot). Correction was done by applying localized regression by locally weighted scatterplot smoothing (LOWESS) (right plot). Note the differently scaled y-axis between the plots.

Afterwards, the normalization of the measurement in the three samples (newly transcribed, pre-existing and total RNA) was performed by linear regression analysis exactly as described before (Doelken *et al.,* 2008). The mRNA half-lives were then calculated based on the normalized ratios given by newly transcribed RNA/total RNA and preexisting RNA/total RNA, according to the formula:

T(1/2)=-t*ln(2)/ln(1-1/(1+Ratio(existing/total)/Ratio(new/total)))

II.21. Calculation of absolute mRNA copy numbers

Absolute cellular mRNA copy numbers were determined based on the measured cellular amount of poly-adenylated mRNA and the number of sequencing reads obtained for the unfractionated mRNA sample (i.e. newly synthesized and preexisting mRNA) similar to the method described previously (Mortazavi *et al.,* 2008). First, the yield of total RNA extracted per 3T3 cell was measured to as 25.5 pg using a NanoDrop ND-1000 UV-VIS spectrophotometer (Thermo Fisher). In three independent experiments obtained 331.92, 304.56 and 360 ng of polyadenylated mRNA from 10 µg total RNA. Therefore, the fraction of mRNA in the cellular RNA pool is 3.3 %, corresponding to 8.5 x 10^{-18} g mRNA per cell. The base composition of the 3T3 mRNAs transcriptome was 0.4 : 0.2 : 0.2 : 0.2 (A:C:G:U) according to the obtained reads. Consequently, the average molecular weight per nucleotide is 323 g/mole. This allows for calculation of the total amount of mRNA nucleotides per cell (T, 2.6x10^{-15} mole/cell). The cellular copy numbers of individual mRNAs (x) can be calculated from the number of sequencing reads mapping uniquely to the transcript (reads_transcript), the total number of reads (reads_total) and the transcript length (L) according to the equation

$$\frac{reads\_transcript}{reads\_total} = \frac{x \times L}{T}$$

For example, the popular "housekeeping" mRNA of GAPDH was identified with 19,392 reads out of 10,854,285 total reads. With a transcript length of 1,232 nucleotides one arrives at 2,270 copies, respectively. A previous study carefully determined GAPDH copy number in NIH3T3 cells by single-cell quantitative real-time PCR or bulk measurements (Marcus *et al.*, 2006). The calculated value falls almost exactly between the previously two published values for bulk and single cell data (1,840 and 2,940, respectively). Note that due to the presence of many GAPDH pseudogenes in the mouse genome all reads mapping to a single GAPDH transcript were counted rather than the corrected uniquely mappable fraction (see below).

II.21.1. Correction for different transcript mappabilities

Correction of different transcript mappabilities was performed by Wei Chen, head of the scientific genomics platform at the Berlin Institute for Medical Systems Biology (BIMSB, Robert-Rössle Str. 10, D-13125 Berlin, Germany):

Absolute mRNA quantification is complicated by the fact that some mRNAs have a high degree of sequence identity to other mRNAs or pseudogenes which results in a smaller percentage of uniquely mappable reads. To correct for this effect the fraction *f* of all the exonic positions on which 36nt sequencing reads that can be uniquely mapped for all mRNAs was determined by *in silico* simulation. Briefly, 36nt Solexa sequencing reads were first simulated from every Refseq transcripts with each position along the transcript (except the last 35 nt) as the first base of one sequencing read. Simulated reads were then mapped to the mouse genome reference sequence and used to estimate transcript abundance using the same procedure as for the real data. The average value of *f* was 72%. mRNAs with $f < 10\%$ were excluded from further analysis. For all other genes, the raw mRNA copy numbers were multiplied by 72 / *f* to correct for different mappability.

II.22. Quantitative real-time (qRT) PCR

To validate RNA-sequencing results used for mRNA half-life calculations, the abundance of 19 randomly selected genes was analyzed via quantitative reverse-transcription polymerase chain reaction. Using the exact same eluate (E) and flow-through (Ft) samples for qRT-PCR as used for RNA-Seq allowed the direct comparison of the obtained E/Ft abundance ratios.

Single-stranded cDNA was generated by reverse transcription of total RNA in a 20 µl reaction volume using SuperScript III Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions. Each 20 µl reaction contained 1.5 µg of total RNA, 250 ng of random hexamer primers (Invitrogen), 1 µl of 10 mM dNTP Mix (10 mM each dATP, dGTP, dCTP and dTTP), 4 µl 5x First Strand Buffer, 2 µl 0.1 M DTT, 1 µl RNasin (Promega), 1 µl of SuperScript III Reverse Transcriptase and RNase-free water. Each gene was run in technical triplicates employing primer pairs targeting mRNAs (cf. Table II.3). Primer pairs were designed using the Primer3

(http://frodo.wi.mit.edu/primer3/) and purchased from BioTez. To eliminate the possibility of genomic DNA amplification all primers were spanning exon-exon boundaries.

qPCR reaction master mix was assembled on ice (25-50 ng cDNA template per reaction; 1 µl of 10 µM forward and reverse primer mix; 2.5 µl 10 x Taq-buffer (Roboklon); 1 µl of 10 mM dNTP Mix; 1.25 µl 50x BSA (NEB), 2.5 µl of 2.5 mM $MgCl_2$ (Roboklon), 1 µl 20x SYBR green (Applied Biosystems), 0.4 µl Perpetual Taq DNA polymerase (2.5 U) (Roboklon); ad. 25 µl ddH_2O). Changes in reporter signal (SYBR green) was recorded in real time with an iQ5 cycler (Bio-Rad). The program was as follows: initial denaturation for 2 min 94°C, followed by 40 cycles of 20 sec at 94°C, 30 sec at 58°C and 40 sec at 72°C. Relative changes in mRNA levels of target genes were computed by normalizing the amount of transcripts to the levels of Ccnd1 (NM_007631) mRNA as an internal control and according the $2^{-\Delta\Delta C_t}$ method (Livak & Schmittgen, 2001).

Table II.3: Primers for qRT-PCR

Gene	Accession ID	Sequence 5´->3´
Bax	NM_007527_FW	ATGCGTCCACCAAGAAGCTGA
Bax	NM_007527_RW	AGCAATCATCCTCTGCAGCTCC
Cav1	NM_007616_FW	ACGTAGACTCCGAGGGACAT
Cav1	NM_007616_RW	CGCGTCATACACTTGCTTCT
Ccnd1	NM_007631_FW	CTGGATGCTGGAGGTCTGT
Ccnd1	NM_007631_RW	AAGCGGTCCAGGTAGTTCAT
Hars	NM_008214_FW	ATCAAGGCTGAGCTGCTGTA
Hars	NM_008214_RW	CTTGAGTTCCTGCTCACCAA
Hsp90ab1	NM_008302_FW	TTTGGTGTCGGATTCTACTCG
Hsp90ab1	NM_008302_RW	CAGACGACTCCCAGGCATAC
Rac1	NM_009007_FW	GAAAGAGATCGGTGCTGTCAA
Rac1	NM_009007_RW	CAACAGCAGGCATTTTCTCTT
Rpl13a	NM_009438_FW	AGCGCCTCAAGGTGTTGGA
Rpl13a	NM_009438_RW	GAGTGGCTGTCACTGCCTGGTA
Fth1	NM_010239_FW	CCTGCAGGATATAAAGAAACCA
Fth1	NM_010239_RW	TAGCCAGTTTGTGCAGTTCC
Elavl1	NM_010485_FW	ACCTCCCTCAGAACATGACC
Elavl1	NM_010485_RW	CCAAGCTGTGTCCTGCTACT
Igf1r	NM_010513_FW	GACTCGGATGGCTTCGTTAT
Igf1r	NM_010513_RW	AGTACATGCTCTGGGTGCTG
Ntan1	NM_010946_FW	TCCAATCATTTATGGCATCG
Ntan1	NM_010946_RW	GCTAATCATTGGGCCTCCT
Tuba1a	NM_011653_FW	AAGGAGGATGCTGCCAATAA
Tuba1a	NM_011653_RW	GCTGTGGAAAACCAAGAAGC
Vim	NM_011701_FW	CAGAGAGAGGAAGCCGAAAG
Vim	NM_011701_RW	GTGCCAGAGAAGCATTGTCA
Ccl5	NM_013653_FW	ATATGGCTCGGACACCACTC
Ccl5	NM_013653_RW	GTGACAAACACGACTGCAAGA
Xpo4	NM_020506_FW	CTGAGCACTGAGGTTCTCCC

Xpo4	NM_020506_RW	ATTCCACGGTAACTGCAAGC
Tbl1x	NM_020601_FW	TTGTGCCTGGAATCCTGTTA
Tbl1x	NM_020601_RW	TAAGGTTCCATATCCTCGCA
Cox6c	NM_053071_FW	GGGAAGGACGTTGGTGTAGA
Cox6c	NM_053071_RW	CTTATAGGCAGCGGCAACTC
Vps4a	NM_126165_FW	CGCCATCAAATATGAAGCAC
Vps4a	NM_126165_RW	GGTACTGCATGCACTTTGCT
Acaca	NM_133360_FW	GAAGTCAGAGCCACGGCACA
Acaca	NM_133360_RW	GGCAATCTCAGTTCAAGCCAGTC
Lmnb1	NM_010721_FW	GGCGGCACTAAACTCTAAGG
Lmnb1	NM_010721_RV	ATGCTTCTAGCTGGGCAATC
Ipo9	NM_153774_FW	GTGGATTCGGATGGAGAAGT
Ipo9	NM_153774_RV	TAGGGCATGGACAAACTCAA

II.23. Cycloheximide-chase analysis

To validate protein stability determined by pulsed SILAC, classic cycloheximide-chase analysis was performed with proteins spanning a reasonable range of protein half-lives (1 - 60 h). For this purpose, protein synthesis of NIH3T3 cells was blocked by adding cycloheximide directly to the culture medium at a final concentration of 50 µg/ml. Cells were harvested after cycloheximide treatment at indicated time points followed by protein extraction, SDS-PAGE and immunoblotting to visualize protein degradation of respective proteins.

II.24. Liquid-chromatography mass spectrometry (LC-MS)

II.24.1. In-gel digestion

To decrease sample complexity, proteins were separated based on their molecular weight by 1-D discontinuous SDS-PAGE prior to digestion (Shevchenko *et al.,* 2006). Cell pellets were lysed in appropriate amounts of RIPA buffer for 20 min on ice. The lysates were cleared by centrifugation for 10 min (14,000 rpm at 4°C) and transferred to fresh tubes. SDS-PAGE was performed with whole-cell lysates using NuPAGE Novex 4 to 12% gradient gels (Invitrogen) under reducing conditions according to the manufacturer's instructions. Proteins were fixed in fixative solution and stained afterwards with the Colloidal Blue staining Kit (Invitrogen). Whole gel lanes were cut into several gel slices (up to 20 slices) and washed three times with ABC buffer to remove Coomassie staining and detergents (SDS). Afterwards, reduction buffer was applied to each slice for 30 min at 56°C for reduction of disulfide bonds followed by alkylation with iodoacetamide (IAA) for 30 min at 37°C in the dark. After another washing step, the gel pieces were dehydrated in 100% ethanol. The gel slices were rehydrated with sequence grade modified trypsin (protein:enzyme ratio of 50:1) resolved in ABC buffer and incubated overnight at 37°C. Protease activity was quenched by adding trifluoroacetic (Uvasol grade, Merck) acid to adjust the pH to < 2. Peptide extraction from the gel matrix was performed by

incubating the gel pieces twice with extraction buffer for 20 min each followed by complete dehydration of the gel pieces with 100% acetonitrile (HPLC gradient grade, Fischer Scientific). After each incubation step, the gel pieces were spun down, supernatants were recovered and finally combined. Lastly, samples were dried in a vacuum concentrator (Eppendorf) until 10-20% original volume to remove the acetonitrile and applied to StageTips (see below) for desalting and further sample concentration.

II.24.2. In-solution digestion

Cells were lysed in appropriate amounts of denaturation buffer for 20 min on ice. The lysates were cleared by centrifugation for 10 min (14,000 rpm at 4°C) and transferred to fresh tubes. Protein samples were reduced for 30 min at RT in 10 mM dithiothreitol (DTT) solution followed by alkylation for 20 min by 55 mM iodacetamide (IAA) in the dark at RT. The endoproteinase LysC (Wako) was added (protein:enzyme ratio of 50:1) and incubated for 4 h at room temperature. After dilution of the sample with four times ABC buffer, sequence grade modified trypsin (Promega; protein:enzyme ratio of 50:1) was used for overnight digestion. Trypsin and Lys-C activity was quenched by adding trifluoroacetic acid to adjust the pH to < 2, and peptides were extracted and desalted using StageTips (see below).

II.24.3. Preparation of Stop and Go Extraction Tips (StageTips)

StageTips (Rappsilber *et al.,* 2003) were assembled by punching out small discs of C18 Empore (3M) filter using a 22 G syringe and ejecting the discs into 200 µl pipet tips. The resulting C18 Empore columns were conditioned by methanol and equilibrated with buffer A. Peptide solutions from in-solution or in-gel digests were adjusted to pH<2.5 and forced through the C18 Empore column by centrifugation at a maximum speed of 6,000 rpm. The columns were washed once with buffer A. Peptides were eluted from the StageTips by applying 60 µl of buffer B to the microcolumns. Peptides were directly eluted into an HPLC autosampler plate and concentrated in a vacuum concentrator (Eppendorf) down to 2-3 µl. After mixing the peptide solutions with an equal amount of buffer A the samples were ready for LC-MS/MS analysis.

II.24.4. HPLC and mass spectrometry

Basic instrumental setup of reversed-phase liquid chromatography (rpHPLC) mass spectrometry is depicted in Fig. II.2. RpHPLC was performed with either the Agilent HPLC 1200 or Eksigent NanoLC – 1D Plus system using self-made fritless C18 microcolumns (Ishihama *et al.*, 2002) (75 m ID packed with ReproSil-Pur C18-AQ 3-µm resin, Dr. Maisch GmbH) connected on-line to the electrospray ion source (Proxeon) of a LTQ-Orbitrap mass spectrometer (classic, XL or Velos instruments, Thermo Fisher). Peptide samples were picked up by the autosamples and loaded onto the column with a flow rate of 500 nl/min (Agilent HPLC 1200) or 250 nl/min (Eksigent HPLC) and subsequently eluted with a flow rate of 200 nl/min with a 10 to 60 % acetonitrile gradient over up to 6 h in 0.5% acetic acid for online MS analysis. The LTQ-Orbitraps were operated in the data dependent mode (DDA) with the full scan in the Orbitrap followed by

consecutive MS/MS scans in the LTQ. The precursor ion scans (m/z 300–1700) were acquired in the Orbitrap part of the instrument (resolution R = 60,000; target value of 1 x 10^6). In parallel, the five (classic, XL) or 20 (Velos) most intense ions were isolated (target value of 5,000 for classic and XL instruments, 3,000 for the Velos instrument; monoisotopic precursor selection enabled) and fragmented in the LTQ part of the instrument by collision induced dissociation (CID; normalized collision energy 35 %; wideband activation enabled). Ions with an unassigned charge state and singly charged ions were rejected. Former target ions selected for MS/MS

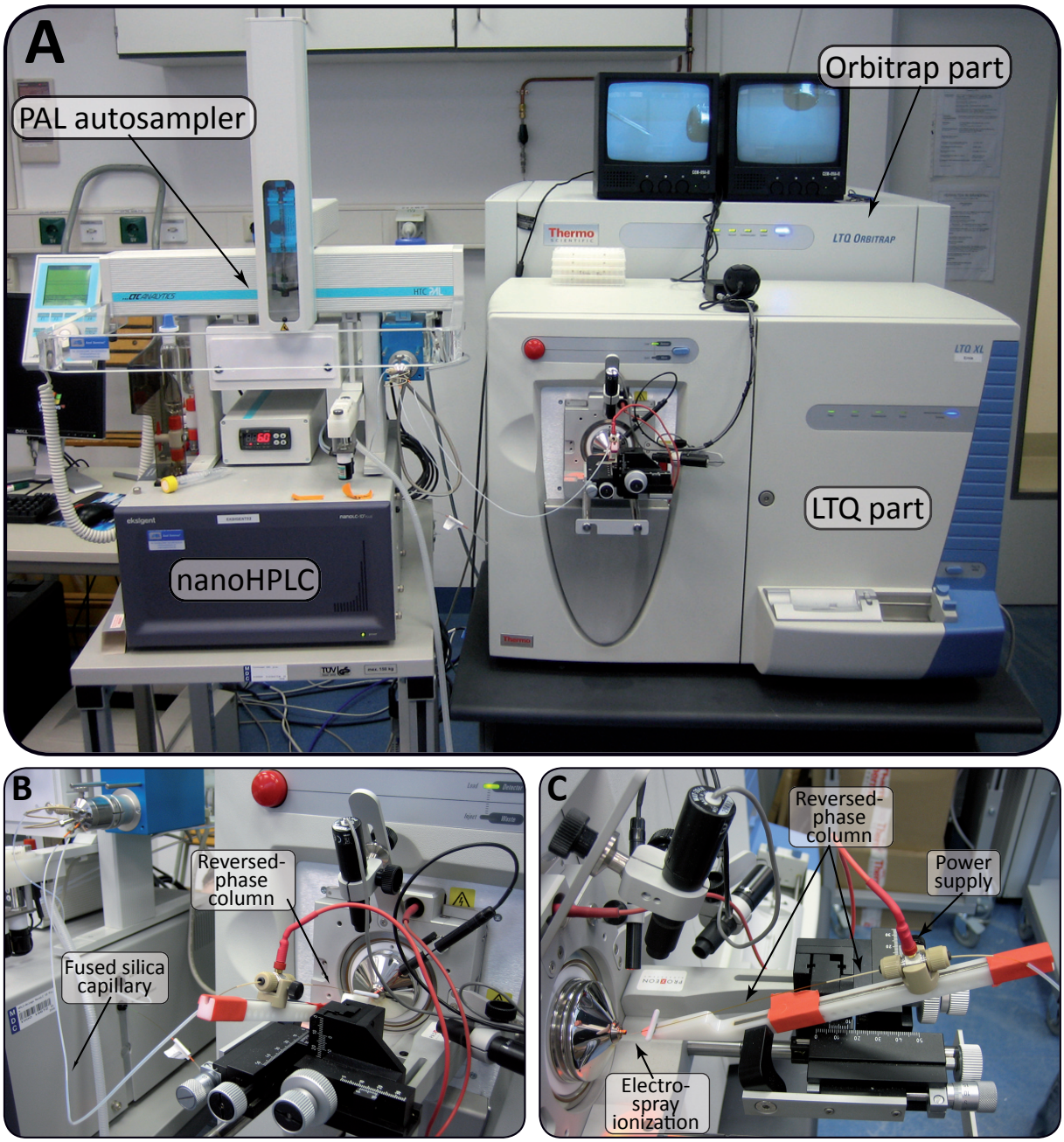


Fig. II.2 | Basic instrumental setup of liquid chromatography-mass spectrometry (LC-MS). (A) Following sample processing, peptide mixtures are stored at 6°C in 96 well plates in the PAL autosampler. The autosampler injects the peptide mixture into a fused silica capillary (B) that is connected to the actual reversed-phase column. Peptides are retained on the column as it is packed with hydrophobic alkyl chain (octadecyl carbon chain, C_{18}) bonded silica constituting the stationary phase. As the fraction of organic solvent (e.g. acetonitrile) in the mobile phase increases, peptides elute sequentially according to their hydrophobicity. Notably, by using nanoHPLC systems flow rates at nanoscale (200 nl/min) are achieved thereby increasing the sensitivity of the mass spectrometer. (C) Peptides eluting from the C_{18} column are immediately ionized in close proximity to the entrance of the mass spectrometer by means of a high electric potential applied to the end of the column, referred to as electrospray ionization (ESI).

were dynamically excluded for 60 s. Total cycle time for one full scan plus up to 20 MS/MS scans was approximately 2 s.

II.24.5. Processing of mass spectrometry data

Identification and quantification of proteins was carried out with the MaxQuant software package (Cox *et al.,* 2008) (version 1.0.7.3 for the pSILAC method development and microRNA target identification; version 1.0.13.13 for half-lives measurements and the establishment of the iBAQ approach). In essence, isotope clusters and SILAC doublets/triplets were extracted, re-calibrated and quantified in the raw data files with Quant.exe (medium labels: Arg6 and Lys4, heavy labels: Arg10 and Lys8; maximum of three labeled amino acids per peptide; polymer detection enabled; top 6 MS/MS peaks per 100 Da). The generated peak lists (msm files) were submitted to a MASCOT search engine (version 2.2, MatrixScience) and searched against an in-house curated concatenated target-decoy database of forward and reversed proteins. Species-specific IPI protein databases (version 3.37 of the IPI human database for the pSILAC method development and microRNA target identification; version 3.64 of the IPI mouse database for protein half-lives measurements and the establishment of the iBAQ approach) and the *E. coli* database (also for the establishment of the iBAQ approach, obtained from: http://cmr.jcvi.org/cgi-bin/CMR/shared/MakeFrontPages.cgi?page=batchdownload) were employed and supplemented with common contaminants (e.g. trypsin, BSA). Databases used in the iBAQ approach were additionally supplemented with UPS2 protein sequences. Full tryptic specificity was required and a maximum of two missed cleavages and a mass tolerance of 0.5 Da for fragment ions was applied. The initial mass accuracy cut-off on the parent ion was 7 ppm but subsequently narrowed down by filtering based on hits to reversed peptides in the target-decoy database. Oxidation of methionine and acetylation of the protein N-terminus were used as variable modifications, carbamidomethylation of cysteine as a fixed modification. Filtering of putative MASCOT peptide identifications, assembly of proteins and re-quantification was performed with Identify.exe (part of MaxQuant). A minimum peptide length of 6 amino acids and a minimum of two peptides per protein group (with at least one of the two being unique in the database) was required. False discovery rates were estimated based on matches to reversed sequences in the concatenated target-decoy database (Elias *et al.,* 2007). A maximum false discovery rate of 1% at both the peptide and the protein level was allowed. Peptides were assigned to protein groups (that is a cluster of a base protein plus additional proteins matching to a subset of the same peptides). Protein groups containing matches to proteins from the reversed database or contaminants were discarded. To quantify changes in protein production (pSILAC method development and microRNA target identification), the median of all H/M peptide ratios using only unique peptides and non-unique peptides assigned to the protein group with the highest number of peptides (“razor” peptides) were calculated. Likewise, *de novo* protein synthesis (measurement of protein half-lives) given by the median of all H/L peptide ratios for a specific protein was quantifiable. For subsequent data analysis, only protein quantifications based on at least three independent ratio measurements were

considered (for protein half-life measurements, at least three peptide ratios at all three time points combined were required).

II.25. Application of the Universal Proteomics Standard (UPS)

The Universal Proteomics Standard (UPS2, Sigma-Aldrich) consists of 48 accurately quantified human proteins in lyophilized form (10.6 µg total). The whole standard was dissolved in 20 µl denaturation buffer and mixed with the samples as follows: UPS2 alone (3.2 µg UPS2), in combination with *E. coli* lysate (2.1 µg UPS2 + 1.5 µg *E. coli* and 1.1 µg UPS2 + 3.0 µg *E. coli*), yeast lysate (4.24 µg UPS2 + 12 µg yeast ~ 6.4 x 10^6 cells) and NIH3T3 lysate (4.24 µg UPS2 + 13 µg NIH3T3 ~ 2.8 x 10^5). All samples were subjected to in-solution digest.

II.26. Calculation of cellular protein copy numbers by intensity-based absolute quantification (iBAQ)

MaxQuant computes protein intensities as the sum of all identified peptide intensities (maximum detector peak intensities of the peptide elution profile, including all peaks in the isotope cluster). These protein intensities were divided by the number of theoretically observable peptides (calculated by *in silico* protein digestion with a PERL script, all fully tryptic peptides between 6 and 30 amino acids were counted while missed cleavages were neglected). The resulting iBAQ intensities were log-transformed and plotted against known log-transformed absolute molar amounts of the spiked-in standard proteins (UPS2 standard). Linear regression was used to fit iBAQ intensities to absolute standard protein amounts. The slope and intercept from this linear regression was then used to convert iBAQ intensities to mole amounts for all identified proteins. Cellular copy numbers were derived by dividing through the number of cells used in the respective experiment. In doing so, for NIH3T3 cells absolute cellular copy numbers for almost 2,000 proteins were obtained. To arrive at cellular copy numbers for the 6,445 proteins identified (in-gel digest, 20 slices) in the half-life measurements, the copy/numbers obtained in the in-solution digest were used to scale the protein intensities from the in-gel digest by linear regression.

II.27. Cluster analysis of gene ontology (GO) terms

In order to test whether specific genes with certain combinations of mRNA and protein stability have distinct biological functions, regimes were defined as follows: First, genes were separately sorted according to their mRNA and protein half-lives. Second, the top third of mRNA half-lives was combined with either the upper or lower third of protein half-lives, and *vice versa* for protein half-lives. This resulted in four regions comprising distinct combinations of mRNA and protein half-lives. Next, gene ontology analysis using the DAVID Bioinformatics Database (DAVID Bioinformatics Resources, http://david.abcc.ncifcrf.gov/) (Huang da *et al.,* 2009) was performed with the IPI protein identifier of genes in each region to find enriched

biological processes. Calculation of over-represented GO terms was done using the entire list of identified proteins as background (threshold count = 2; EASE score = 1). Terms with a p-value <0.01 in at least one regime were selected, log- and z-transformed, hierarchically clustered and plotted as a heatmap using an in-house Perl and R script.

II.28. Statistical analysis

Statistical data analysis and linear regression was done using the R project for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria) or Prism 4.0 (GraphPad Software, San Diego, CA, USA).

II.29. Determination of protein half-lives, response times and mathematical modeling of gene expression

All the maths required for derivation of the following equations to quantitatively describe the dynamic properties of gene expression was contributed by Dorothea Busse, scientist in the group of Jana Wolf termed "Mathematical Modelling of Cellular Processes" at the Max-Delbrück-Centrum for Molecular Medicine (MDC, Robert-Rössle Str. 10, D-13125 Berlin, Germany):

II.29.1. Calculation of protein half-lives

SILAC enables the determination of protein turnover by measuring the ratio (r) of protein labeled with heavy amino acids (P_H) and light amino acids (P_L) at different time points (1.5 h, 4.5 h and 13.5 h):

[E1]

$$r = \frac{P_H}{P_L}$$

Proteins labeled with light amino acids (P_L) are assumed to decay exponentially with the degradation rate constant (k_{dp}):

[E2]

$$P_L = P_0 e^{-k_{dp}t}$$

Protein labeled with heavy amino acids (P_H) can be expressed as the difference between total number of a specific protein (P_{total}) and P_L. The total number (P_{total}) is assumed to double during the duration of one cell cycle:

[E3]

$$P_H(t) = P_{total}(t) - P_L(t) = P_0 2^{t/t_{cc}} - P_L(t)$$

P_0 is the protein copy number at the onset of the experiment (t = 0) and t_{cc} is the duration

of one cell cycle. The rate constant of the protein decay (k_{dp}) can then be obtained by linear regression using equations E1-E3:

[E4]

$$k_{dp} = \frac{\sum_{i=1}^{m} \log_e (r_{t_i} + 1) t_i}{\sum_{i=1}^{m} t_i^2} - \frac{\log_e 2}{t_{cc}}$$

where m is the number of time points (t_i) considered and the ratio of the light and heavy amino acids containing fraction of a specific protein (equation E1) at each time point. The half-life of a protein ($T_{1/2}$) is then given by:

[E5]

$$T_{1/2} = \frac{\log_e 2}{k_{dp}}$$

Of all identified proteins (n = 6,445) only those were taken which were counted at least three times by the mass spectrometer regardless at how many time points. As a quality test of linear regression, the coefficient of determination (R^2) was calculated from the correlation coefficient between measured and estimated ratios (two or three time point measurements, cf. Fig. II.16 A and B). Half-lives were only calculated for fits that satisfy $R^2 \geq 0.9$ (n = 4,906) or measurements of one time point with a minimum three counts (n = 95). For very unstable proteins the ratio measurements saturated at the late time point. Therefore, for proteins that failed the quality check ($R^2 < 0.9$) but had a ratio $r \geq 0.5$ at the first time point (1.5 h), only this time point was used to calculate the half-life.

II.29.2. Quantitative model of gene expression

A minimal description of the dynamics of transcription and translation includes the synthesis and degradation of mRNA and protein only:

$$\frac{dR}{dt} = v_{sr} - k_{dr} R$$
$$\frac{dP}{dt} = k_{sp} R - k_{dp} P$$

The mRNA *(R)* is synthesized with rate v_{sr} and degraded proportional to the number of mRNAs with the rate constant k_{dr}. Proteins *(P)* are synthesized depending on the number of mRNA molecules and the translation rate constant (k_{sp}). Protein degradation is determined by k_{dp} and the protein level. This system of ordinary differential equation can be solved analytically. The experimental data introduced in this thesis are mean mRNA and protein levels in a population of growing non-synchronous cells. Assuming a homogenous distribution of cells over the duration of one cell cycle (t_{cc}), mean mRNA and protein levels of the population are equal to

the averaged copy numbers of one cell cycle

$$\langle X \rangle = \frac{1}{t_{cc}} \int_0^{t_{cc}} X(t)dt \,,$$

where *X* represents either mRNA *(R)* or protein *(P)* copy numbers. The dynamics of mRNA (*R(t)*) and protein (*P(t)*) is given by the solution of ordinary differential equation system described above. The initial conditions of mRNA and protein copy numbers are given by $R(0) = R_0$ and $P_{(0)} = P_0$, respectively. The copy numbers are assumed to double over one cell cycle to balance the distribution of mRNA and protein between the two daughter cells after cell division, therefore $X(t_{cc}) = 2X_0$. Solving this equation for mRNA and protein allows describing the relationship between mean cellular mRNA and protein levels, synthesis and degradation rates.

II.29.3. Response times

The above presented model of quantitative gene expression (cf. II.29.2) can also be used to calculate characteristic response times τ that describe how fast protein levels relax towards the steady-state after perturbation. Response times were first defined and used to characterize the response to flux changes in metabolic networks (Heinrich & Rapoport 1975; Cascante *et al.*, 1995), and more recently adapted to signaling networks (Heinrich *et al.*, 2002; Klipp, 2009). A plain description of the characteristic time for the linear gene expression model is given by:

$$\tau_{protein} = \frac{\int_0^{\infty} P_{ss} - P(t)dt}{P_{ss} - P_0}$$

Here the response time ($\tau_{protein}$) characterizes the approximation of the protein level in the steady state ($P_{ss} = (v_{sr}k_{sp})/(k_{dr}k_{dp})$). The protein kinetics (*P(t)*) is again determined by the solution of the ordinary differntial equation system. The response time ($\tau_{protein}$) is then given by

$$\tau_{protein} = \frac{1}{k_{dp}} + \frac{k_{sp}}{k_{dr}k_{dp}} \frac{R_{ss} - R_0}{P_{ss} - P_0}$$

where k_{dr} and k_{dp} are the degradation rate constants of mRNA and protein, respectively. The parameter k_{sp} is the translation rate constant and R_0, R_{ss} and P_0, P_{ss} are the initial and steady-state levels of mRNA and protein, respectively. Assuming (i) a perturbation which affects both mRNA and protein levels and (ii) steady-state conditions before and after the perturbation, the response time only depends on the degradation rate constants of mRNA (k_{dr}) and protein (k_{dp}):

$$\tau_{protein} = \frac{1}{k_{dr}} + \frac{1}{k_{dp}}$$

II.30. Energy calculations

With the quantitiative model (cf. II.29.2) the estimation of the theoretical energy costs (in high energy phosphate bonds) to synthesize mRNAs (i.e. primary transcripts) and proteins from their building blocks (nucleotide monophosphates and amino acids, respectively) was possible. For transcription, two ATP molecules are needed for adding a single nucleotide (elongation) plus another two ATPs for the initiation step (Voet, 1995). For protein synthesis, four ATP molecules are need for the addition of single amino acids to the nascent chain (2 ATPs for tRNA loading and 2 ATPs for elongation) plus another 2 ATPs for initiation and one for termination (Moldave, 1995). By multiplication of the above mentioned costs per building block (nucleotide or amino acid) with transcription/translation rates and mRNA/protein lengths overall energy expenses for mRNA and protein synthesis were obtained with:

mRNA synthesis = (transcription rate * primary transcript lengths * 2 ATP) + 2 ATP

protein synthesis = (translation rate * protein length * 4 ATP) + 3 ATP

II.31. Structural features affecting mRNA and protein stability

3'UTR, 5´UTR and CDS sequence lengths, AU-rich elements and Pum2 binding sites

Mouse 3'UTR, 5´UTR and CDS sequences based on the mouse the reference sequence (NCBI Build 37/mm9) were extracted from UCSC Genome Browser (http://genome.ucsc.edu) and the lengths were counted. 3´UTR sequences were searched for AU-rich core motifs ("ATTTA"; Barreau *et al.,* 2006) and Pumilio2 binding sites ("TGTA(A|T|C|G)ATA" with "I" referring to "or"; Hafner *et al.,* 2010) using an in-house Perl script. The average number of either motif in bins of 100 genes were plotted against averaged mRNA half-lives of the same bins. To account for stochastic motif occurences due to vastly different 3´UTR lengths, motif densities (motifs/nucleotide) were calculated by dividing the number of either motif counted per 3´UTR by 3´UTR sequence length.

Correlation of amino acid sequence frequency and protein stability

A possible relationship between the presence of certain amino acids and protein stability was examined as follows: First, relative fractions of amino acids per protein were calculated. Second, relative amino acid fractions were plotted against protein half-lives in bins of 100 genes. Third, Spearman Rank correlation coefficients (R_s) were used to assess the statistical significance of possible correlations.

Correlation of the degree of protein disorder with protein stability

As a proxy for a protein´s "unstructuredness", a publicly available mouse protein database (http://bioinfadmin.cs.ucl.ac.uk/disodb/) was used that predicts potentially disordered amin acid residues in a context-specifc manner and also assigns confidence values accordingly. The overall percentage of potentially disordered amino acids was extracted from the database for individual proteins. A possible correlation was investigated by plotting averaged "unstructuredness" in bins of 100 genes against averaged protein half-lives of the same bins.

III Results

III.1. Global analysis of cellular protein translation by pulsed SILAC (pSILAC)

DNA-microarrays have become a powerful means to routinely study qualitative and quantitative changes of the transcriptome. However, microarrays cannot be employed to study changes occurring at the level of translation. While translational regulation is therefore investigated on a gene by gene basis with radioactive amino acids (e.g. ^{35}S-methionione; Pollard, 1996) or by expression of artificial reporter constructs, e.g. luciferase reporter assays (Dyer *et al.,* 2000), tools for a system-wide investigation of protein translation are lacking.

Mass spectrometry-based proteomics can identify hundreds of proteins in complex mixtures and recently developed methodologies offer the opportunity to obtain quantitative proteomic information. For instance, the SILAC approach can highlight differences in protein composition between samples and has also recently been adapted to measure protein turnover in model organisms like yeast but also in vertebrates (Doherty *et al.,* 2005; Pratt *et al.,* 2002; Milner *et al.,* 2006). In this case, the organism of interest is pulse labeled with heavy amino acids for a certain period of time. The ratio of heavy (H) to light (L) peptides indicates the turnover rate of the respective protein. However, as protein turnover is affected by both synthesis and degradation the H/L ratio does not directly provide information about the translation rate. For instance, a high H/L ratio could either indicate a high translation rate of a stable protein or a low translation rate of a protein that is rapidly degraded.

III.1.1. Establishment of a pulsed SILAC approach with two heavy-stable isotopes

As SILAC metabolically labels exclusively newly-synthesized proteins, a direct comparison of translation rates between two samples should be achievable by introduction of a second heavy-stable isotope label. The concept of this pulsed SILAC (pSILAC) strategy is depicted in Fig. III.1. Cells are first cultivated in standard growth medium with the natural light (L) amino acids. Concomitantly with differential treatment (e.g. application of a certain chemical), cells are transferred to culture medium containing heavy (H) or medium-heavy (M) amino acids (cf. Material & Methods II.4 and II.5). During the following incubation phase all newly synthesized proteins will appear in the H or M form, respectively. Subsequently, both samples are combined and analyzed together by mass spectrometry. In contrast to classic SILAC (cf. Fig. I.4), ionized peptides now emerge as triplicates according to their respective label in the mass spectrum. The abundance ratio of the H versus the M peptide form now exclusively reflects differences in translation of newly-synthesized proteins resulting from a differential treatment. Pre-existing protein amounts, that is, proteins present in the sample before the treatment remain in the

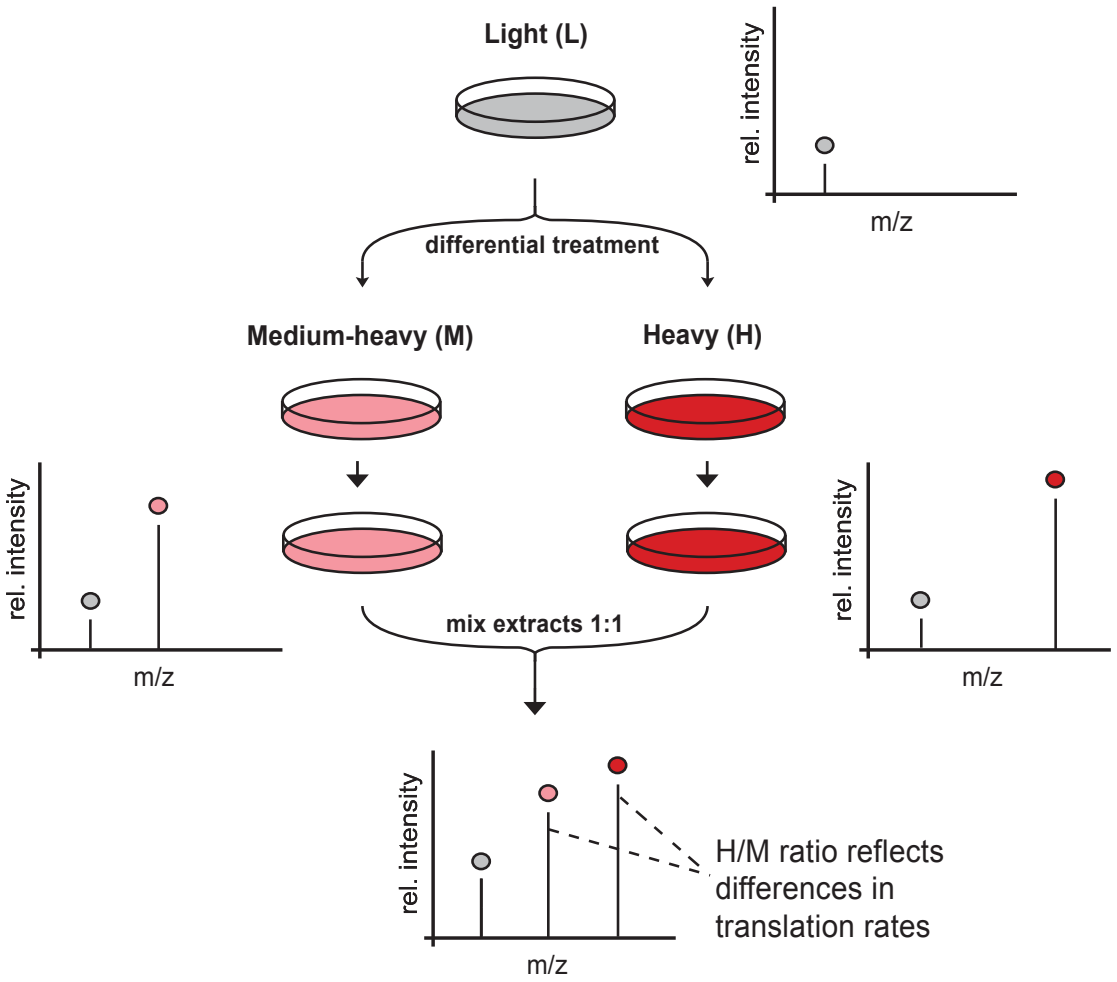


Fig. III.1 | Principle of the pulsed SILAC approach. Concurrently with a differential treatment, cells cultivated in light (L) medium are either transferred to heavy (H) or medium-heavy (M) SILAC medium. From this time point on cells are pulse labeled, that is, newly-synthesized proteins incorporate either the H or the M amino acids. After the labeling phase the two cell populations are mixed and analyzed by mass spectrometry. The ratio of peak intensities of H versus M peptides reflects differences in translation of the corresponding proteins.

L form and can be effectively ignored. The traditional SILAC approach seeks to completely replace light amino acids by their heavy labeled counterparts throughout the proteome. Given that proteins exhibit vastly different turnover rates, changes in gene expression induced by any kind of treatment will affect steady-state protein levels with different kinetics and of different magnitude. Since the pSILAC approach exclusively accounts for differences in protein production, the method is independent of differences in protein stability that would otherwise interfere with accurate quantification. Degradation will to some extent also affect newly-synthesized proteins, however, this degradation will occur for proteins in both the M and the H form to the same degree and will therefore not generally affect differences in translation rates given by H/M ratios.

III.1.2. pSILAC - as good as the traditionals?

In order to evaluate the quantitative parameters of pSILAC, the method was first compared to the well established luciferase assay traditionally used to measure translational regulation. More precisely, a stable HeLa cell line with a firefly luciferase gene under control of pTET (cf. Material

& Methods II.4.1), a promoter responsive to tetracycline controlled transactivators (Gossen & Bujard, 1992) was employed as follows. The kinetic of luciferase induction by treating cells with doxycycline (a tetracycline derivative) in M medium for different lengths of time (1 h, 2 h, 4 h, 6 h, 8 h, 10 h) has been monitored (cf. Material & Methods II.6). In parallel, control cells were treated with doxycycline for 10 h in H medium which corresponds to maximal luciferase induction. Following cell lysis, cells in M medium (treated for variable lengths of times) were mixed in a 1:1 ratio with the control lysate (treated for 10 h) and proteins were separated by SDS-PAGE (cf. Material & Methods II.6). Gel slices in the expected molecular weight range of the luciferase protein were excised, reduced, alkylated and trypsin digested (cf. Material & Methods II.24.1). Extracted peptides were analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) on a high performance mass spectrometer (cf. Material & Methods II.24.4). Despite of more than 500 proteins commonly found in excised gel slices, unequivocal identification of several luciferase peptides for each time point was achieved (cf. Supplementary Table VII.1).

In Fig. III.2 A and B two typical mass spectra of a luciferase derived peptide obtained with the pSILAC approach are depicted. The increasing peptide intensity of the M form over time allowed for quantification of *de novo* translation of the luciferase protein relative to the control

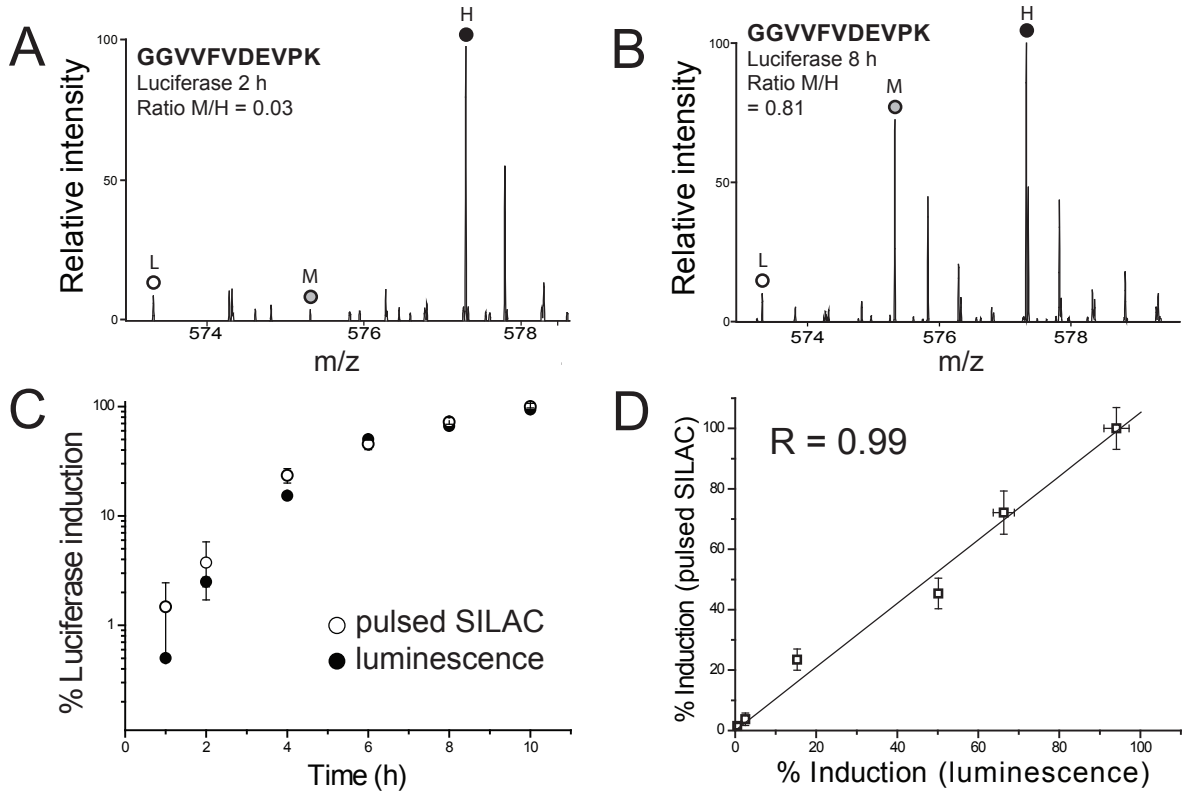


Fig. III.2 | Evaluation of the pulsed SILAC approach. (A, B) Mass spectra of a luciferase-derived peptide (GGVVFVDEVPK) at 2 and 8 h post induction. Luciferase expression was induced by treating transgenic HeLa cells with doxycycline in M medium for different lengths of time. Cells treated for 10 h in H medium were used as an internal reference. The increase in the M form relative to the H form is apparent. Of note, small amounts of luciferase-derived peptides in the light form were also detected. (C) Time course of luciferase induction relative to the reference as measured by pSILAC and luminescence readouts. (D) Plotting pulsed SILAC-based quantification against luminescence-based quantification shows that both methods yielded very similar results. Error bars represent standard deviations calculated by comparing three luminescence measurements or by quantifying different luciferase peptides (pSILAC).

in the H form. In parallel, luciferase expression using the standard luminescence approach was analyzed in the same samples. Here, luminescence-based relative quantification of luciferase induction was performed by referring the absolute luminescence intensities for the different incubation times (M samples) to the control cells (H sample) (cf. Material & Methods II.6). Notably, a small amount of luciferase-derived peptides in the L form (cf. Fig. III.2 A, B) was also detected. As luciferase was not expressed while cells were cultivated in the light medium this peak most likely originates from recycling of light amino acids or the use of light amino acids still present in the cellular precursor pool after transfer to M or H medium.

Direct comparison of results obtained by pSILAC and luminescence readout is shown in Fig. III.2 C. Overall, pSILAC based quantification was found to be very similar to the highly sensitive luciferase assay. Standard deviations given by error bars for SILAC quantification became smaller over time. This observation can likely be attributed to the fact that stable-isotope based quantification works best with peptide ions displaying proper signal-to-noise ratios. More specifically, while decent amounts of newly synthesized luciferase 8 h post-induction made identification of the M peptide peak relatively straightforward (cf. Fig. III.2 B), unambiguous peak detection was more challenging at 2 h post-induction (cf. Fig III.2 A). Consequently, H/M peptide ratios obtained at very early time points exhibit higher deviations than at later time points.

Plotting SILAC quantification against luminescence-based quantification demonstrated excellent overall correlation (Pearson R=0.99, cf. Fig III.2 D). Linearity was observed for abundance ratios between 0.015 and 1, corresponding to 1.5 % and 100 % induction of luciferase expression, respectively. Hence, pSILAC can accurately determine differences in translation rates over at least two orders of magnitude in complex protein mixtures provided that a mass spectrometer with a sufficiently high dynamic range is employed.

III.1.3. pSILAC accurately quantifies over a high range of expression levels

In order to validate that pSILAC accurately quantifies also other proteins than luciferase, experiments with SILAC labeled yeast as a model system were conducted. Yeast cells (strain BY4742 lys2:kanMX4) were completely SILAC labeled in the traditional way by cultivating them over night in minimal media supplemented with light (L), medium-heavy (M) or heavy lysine (H) (cf. Material & Methods II.5 and Table II.1). To simulate different translation rates (i.e. different H/M ratios) the labeled yeast populations were mixed in four different L:M:H ratios (0.5:1:1, 0.5:1:2, 0.5:1:4, 0.5:1:8) spanning a reasonable range of variations in translation. The mixtures were subjected to in-solution digests (cf. Material & Methods II.24.2) and analyzed via mass spectrometry (cf. Material & Methods II.24.4). In each sample, several hundred proteins in all three forms could readily be detected. Log2 fold changes of the obtained H/M ratios were plotted as a function of summed peptide intensities (cf. Fig. III.3). The analysis demonstrated that pSILAC can accurately measure H/M ratios for hundreds of proteins in parallel. Even large deviations in the amount of the H compared to the M form (e.g. an H/M ratio of 8:1) simulating

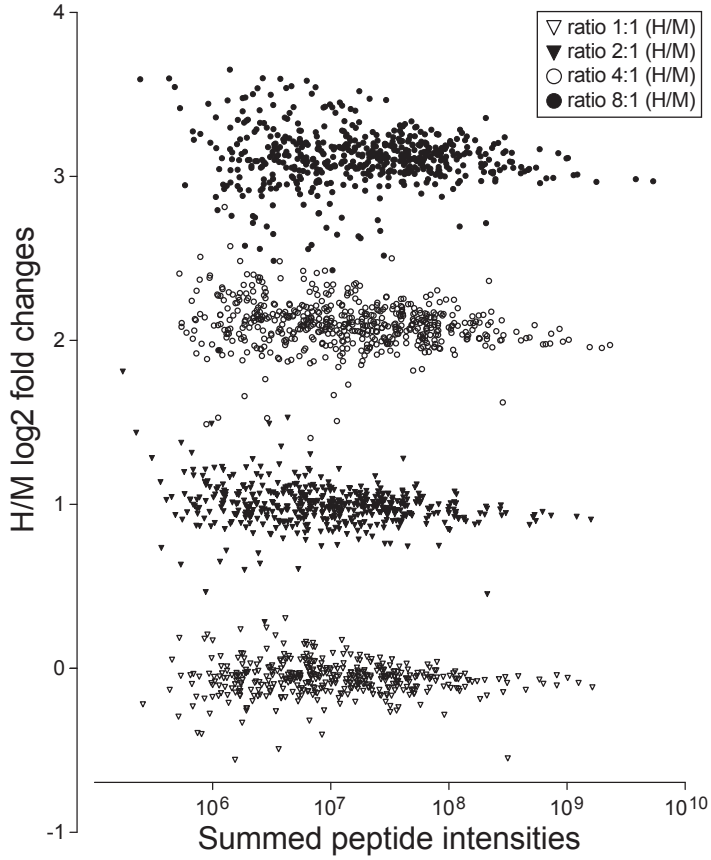


Fig. III.3 | Benchmark test of pSILAC using yeast as a model system. L, M and H labeled yeast cells were mixed in different ratios and analyzed. Log2 fold changes of H versus M proteins were plotted as a function of summed peptide intensities. Note the marked correlation between the measured values and the mixing ratios.

vast differences in translation rates centered closely around the expected log2 fold change of three. As expected, quantification was more accurate for the more highly expressed proteins, that is, proteins represented by more peptides with higher peak intensities. However, also less abundant proteins rarely deviated more than 0.3 log2 fold change from the true value indicating that pSILAC can accurately quantify protein changes over a high range of expression levels.

III.1.4. Applying pSILAC to iron homoeastasis - a proof of principle

Next, the pSILAC approach was employed to study translation in a biologically relevant system. Cellular iron homeostasis has been found to be an ideal model for translational regulation as it has been extensively studied for many years (HENTZE *et al.*, 2004). Here, iron regulatory proteins (IRPs) interact with iron-responsive elements (IREs) which are conserved *cis*-acting hairpin structures found in untranslated regions (UTRs) of mRNAs of iron-related proteins. Binding of IRPs to single IREs in the 5' UTR of mRNAs encoding ferritin heavy (FTH1) and light (FTL) chains is known to block translation of these iron storage proteins. In contrast, interaction between IRPs and multiple IREs in the 3' UTR of the mRNA encoding for the transferrin receptor (TFRC) protects this message from degradation. Cellular iron inhibits the interaction between IRPs and IREs or targets IRPs for degradation. Consequently, iron regulates protein synthesis either by altering mRNA stability or by directly interfering with translation initiation. To investigate the global effect of iron availability on protein synthesis rates, HeLa cells were washed with PBS and transferred to H SILAC medium with simultaneous addition of ferric ammonium citrate (FAC) to a final concentration of 100 µM (cf. Material & Methods II.7). In parallel, control cells were washed and transferred to M SILAC medium without additional iron. After 14 h of incubation, FAC-treated and untreated control cells were harvested, combined and analyzed by LC-MS/

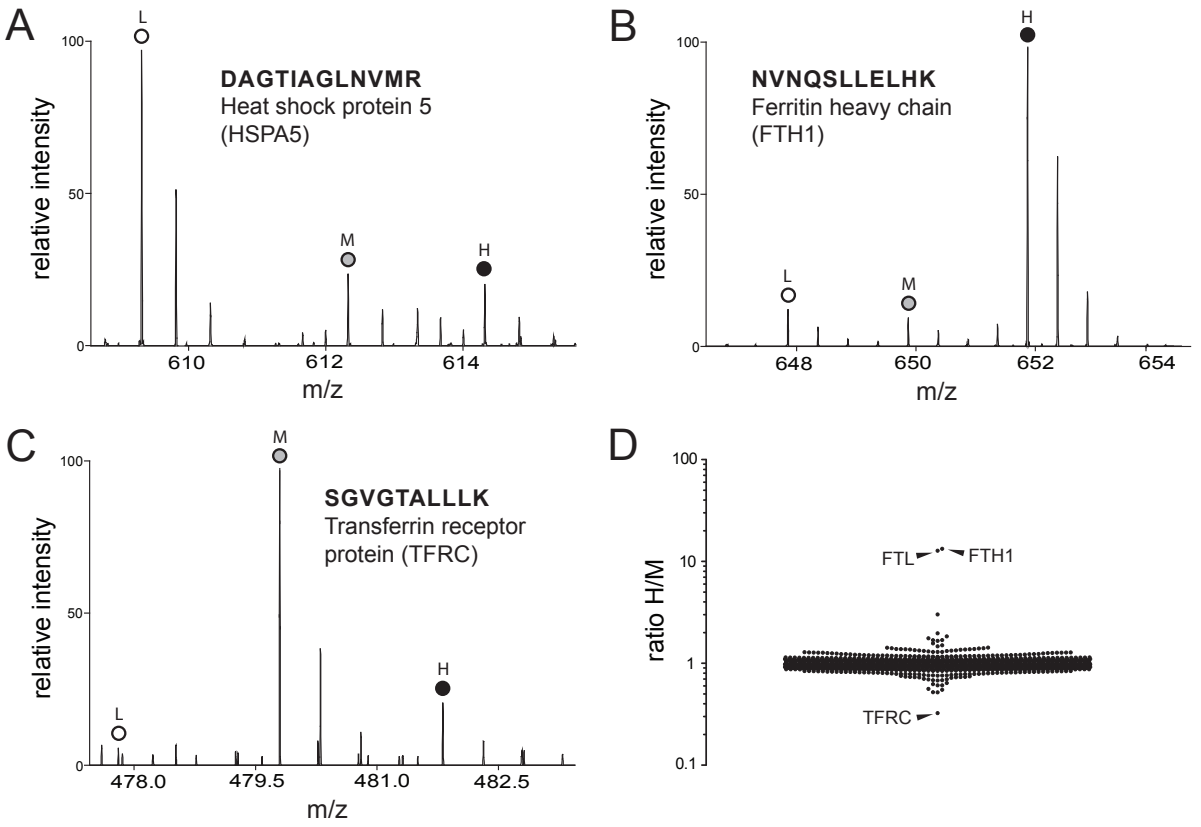


Fig. III.4 | Changes in cellular protein translation rates in response to iron. HeLa cells were treated with iron in H medium or transferred to M medium without additional iron. (A) Translation rates of most proteins were not altered as shown by similar intensities of the H and the M peak. (B) Synthesis of the ferritin heavy (FTH1) and light chains (FTL) was strongly stimulated by iron. (C) Translation of the transferrin receptor protein (TFRC) was repressed by iron treatment. (D) H/M ratios of 1,311 identified proteins show that FTH1, FTL and TFRC have the highest fold changes. Protein synthesis of these proteins was significantly up- or down-regulated by iron ($p < 0.01$, Grubbs' outlier test).

MS. In total, 1,311 proteins could be identified with extremely high confidence (cf. Material & Methods II.24.5).

Exemplary for the vast majority of proteins detected with H:M ratios around 1:1, indicating that the translation of those has not been significantly affected by iron (cf. Fig.III.4 D), the spectrum of a peptide derived from the Heat shock protein 5 (HSPA5) is shown in Fig. III.4 A (cf. Supplementary Table VII.3 for additional examples). As opposed to this, ferritins were found to be 13 times (log2 fold change of ~3.7) more abundant in the H than in the M form, demonstrating that translation of these iron storage proteins was strongly activated (cf. Fig. III.4 B and Supplementary Table VII.2). Consistent with the aforementioned model of iron regulation, a significant down-regulation of the transferrin receptor in response to iron addition could be detected (cf. Fig. III.4 C and Supplementary Table VII.2).

In order to validate these findings, experiments with different incubation times with iron and a complementary approach with the iron chelator deferoxamine (DFO) were performed. All experiments confirmed technical robustness and reproducibility of the pulsed SILAC method. Consistently, in all five biological experiments the ferritins and the transferrin receptor were found to be among the most strongly regulated proteins (cf. Fig. III.5), indicating that pulsed SILAC can accurately identify proteins regulated at the post-transcriptional level.

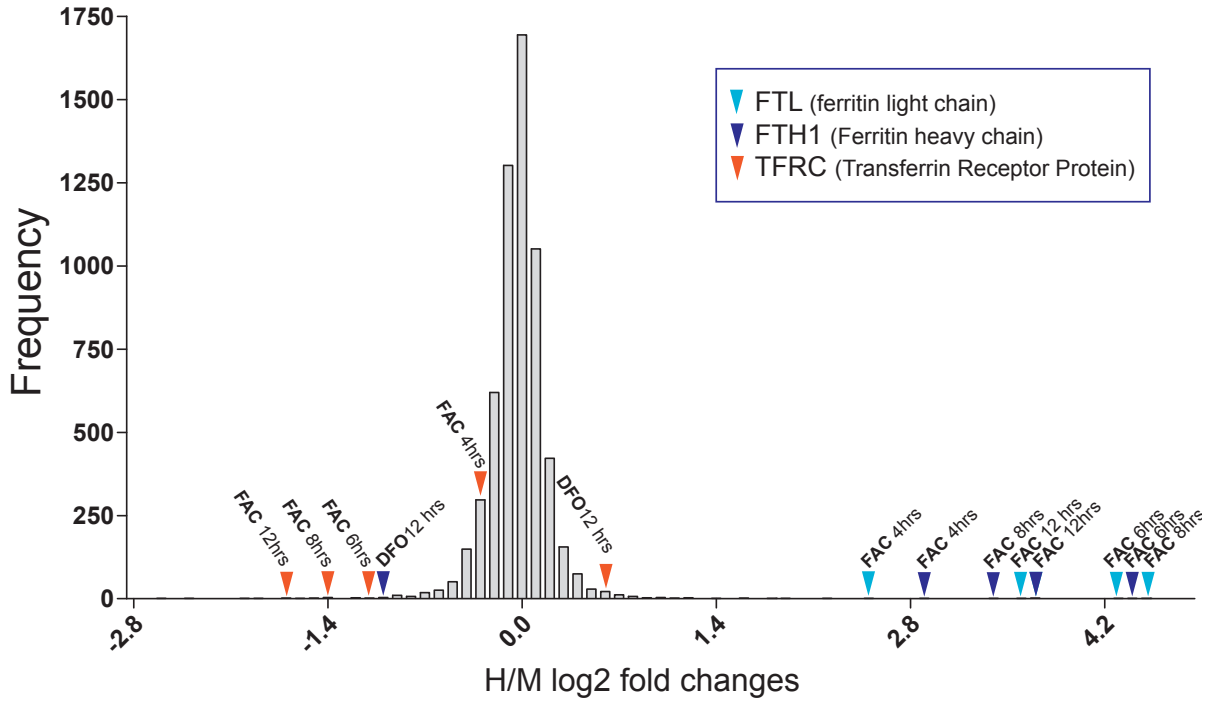


Fig. III.5 | Summary of changes in protein production in HeLa cells in response to iron derived from five independent pSILAC experiments. Cells were either treated with 100 µM ferric ammonium citrate (FAC) for different lengths of time (4, 6, 8 and 12 h) or with the iron chelator deferoxamine (DFO) for 12 h (cf. Material & Methods II.7). Changes in production of all identified proteins relative to untreated controls are plotted as a histogram of log2 fold changes of H/M ratios. Production of most proteins was not significantly affected by iron as seen by the huge peak in the histogram at a log2 fold change of 0. In contrast, ferritin light (FTL) and heavy (FTH1) chain, two proteins known to be translationally regulated by iron availability, were strongly up-regulated by addition of iron at all four time points, while the transferrin receptor (TFRC) was consistently down-regulated. Note that removing iron with DFO resulted in down-regulation of FTH1 and up-regulation of TFRC.

III.2. Widespread changes in protein synthesis induced by microRNAs

As mentioned in the introduction, quantitative approaches to monitor global protein expression changes upon perturbations are now available. However, since microRNAs act at the post-transcriptional level, that is, by affecting mRNA stability and/or mRNA translation, changes in steady-state protein levels upon miRNA over-expression or depletion are not ideally suited to detect their impact on the proteome. Specifically, due to individual protein turnover times, the kinetics with wich microRNA induced changes become apparent at steady-state protein levels may differ substantially.

III.2.1. pSILAC quantifies changes in protein production induced by miRNAs

To overcome these problems, the mass spectrometry-based quantitative pSILAC approach (cf. III.1) ideally fills in this technical gap. To briefly recap, pSILAC employs different stable-isotope labeled amino acids to metabolically pulse-label differentially treated cell populations. Global differences in protein production integrated over the measurement time after the pulse can then be detected (SCHWANHAEUSSER *et al.,* 2009).

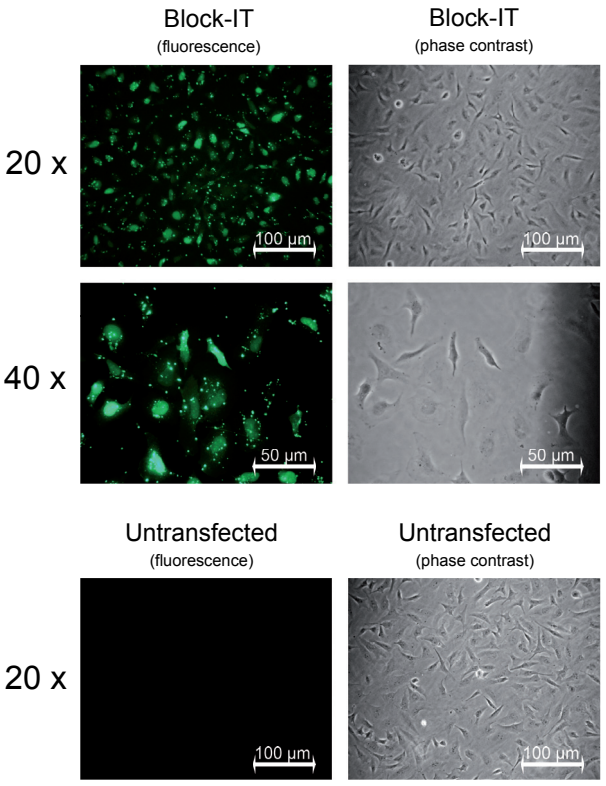


Fig. III.6 | Analysis of transfection efficiency. HeLa cells were transfected with fluorophore-conjugated dsRNA termed Block-IT according to the miRNA transfection protocol (cf. Material & Methods II.8) (left panel). Phase contrast images are presented as a reference (right panel).

In 2005 Johnson and co-workers (LIM *et al.* 2005) demonstrated that ectopic over-expression of single microRNAs results in widespread, mostly mild down-regulation of hundreds of mRNAs. To assess the impact of microRNAs onto the proteome with pSILAC, five miRNAs were individually over-expressed in HeLa cells (cf. Material & Methods II.8). These miRNAs are tissue-specific and either virtually absent in HeLa cells (miR-1, miR-155) or appreciably expressed in various tissues (miR-16, miR-30a, let-7b) including HeLa cells (LANDGRAF *et al.,* 2007).

To check for efficient transfection of small RNAs, HeLa cells were first transfected with fluorophore-conjugated dsRNA (cf. Fig. III.6) according to the applied miRNA transfection protocol (cf. Material & Methods II.9). Fluoresence microscopy confirmed successful delivery of small

RNAs in more than 90 % of all cells.

To examine changes in protein production upon over-expression of individual miRNAs, cells were pulse-labeled in medium-heavy (M) SILAC medium 8 h post-transfection (cf. Fig. III.7). In parallel, MOCK-transfected control cells ("water control") were cultivated in heavy (H) SILAC medium (cf. Material & Methods II.8). Following a labeling time of 24 h, differentially treated HeLa cells were combined and subjected to mass spectrometry analysis. Representative peptide mass spectra demonstrating pSILAC-based quantification of changes in *de novo* synthesis of three proteins upon miRNA-1 or miRNA-155 transfection are depicted in Fig. III.8 A,B and C. Changes in protein synthesis were revealed by the relative intensities of the H and the M peak (H/M ratio). While protein production of β-actin was not significantly altered, indicated by H/M ratio of 1.1, the hepatocyte growth factor receptor (c-MET) displayed down-regulation in protein synthesis upon miRNA-1 over-expression. Likewise, protein production of the CCAAT-enhancer-binding proteins beta (CEBPβ) was substantially repressed following miRNA-155 transfection. Overall, for each of the five miRNAs investigated, changes in protein production of approximately 3,000 to 3,500 proteins were quantified with high confidence (all results for individual microRNAs are freely accessible via web-based database search at http://pSILAC.mdc-berlin.de. Alternatively, all data sets are available as flat-file for download).

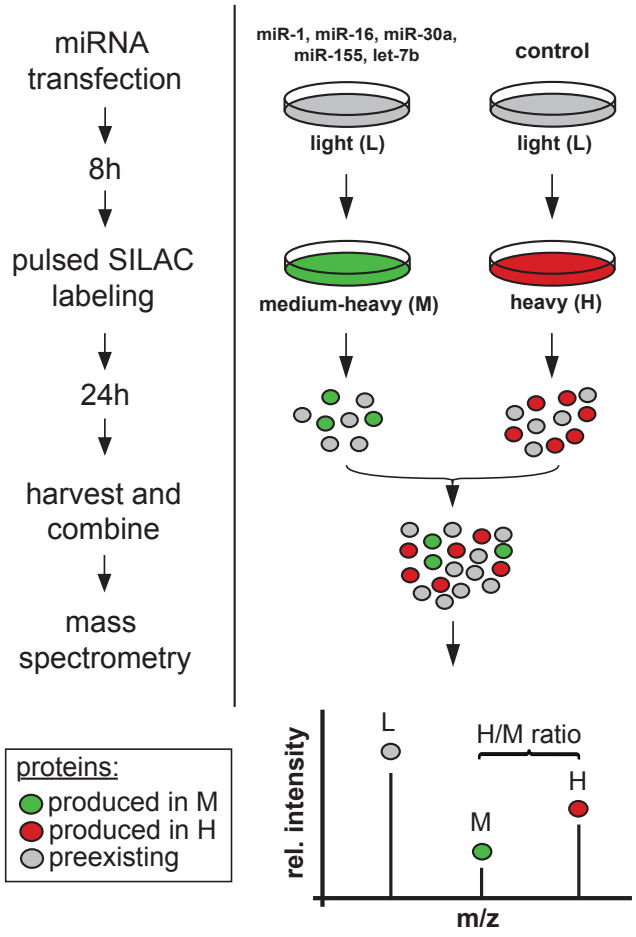


Fig. III.7 | Global analysis of changes in protein production induced by microRNAs. HeLa cells cultivated in normal, .i.e. light (L) medium were either transfected with a miRNA or MOCK transfected. After 8 h, transfected and control cells were pulse labeled by transferring them to culture medium containing medium-heavy (M) or heavy (H) isotope-labeled amino acids, respectively. Newly synthesized proteins will appear in the H or M form. 24 h post-transfection, samples were combined and analyzed by mass spectrometry. Intensity peak ratios between heavy and medium-heavy peptides (H/M ratio) reflect changes in protein production.

To account for biological reproducibility of SILAC measurements, exemplary for miRNA-1 two independent transfection experiments were performed (cf. Material & Methods II.8). As shown in Fig. III. 8 D, the biological replicates showed high correlation over the entire dynamic range indicated by a Pearson correlation coefficient (R) of ~ 0.9. In order to validate at least some targets found with repressed protein production by the pSILAC approach, altogether 16 proteins were independently reexamined with conventional Western blotting (cf. Material & Methods II.14 and Supplementary Fig. VII.1). Though Western blotting shows changes in protein steady-state levels and is hence potentially compromised by individual protein turnover, results did in all cases qualitatively validate the pSILAC measurements, i.e. overall trends apparent from the blots agreed well with the pSILAC results.

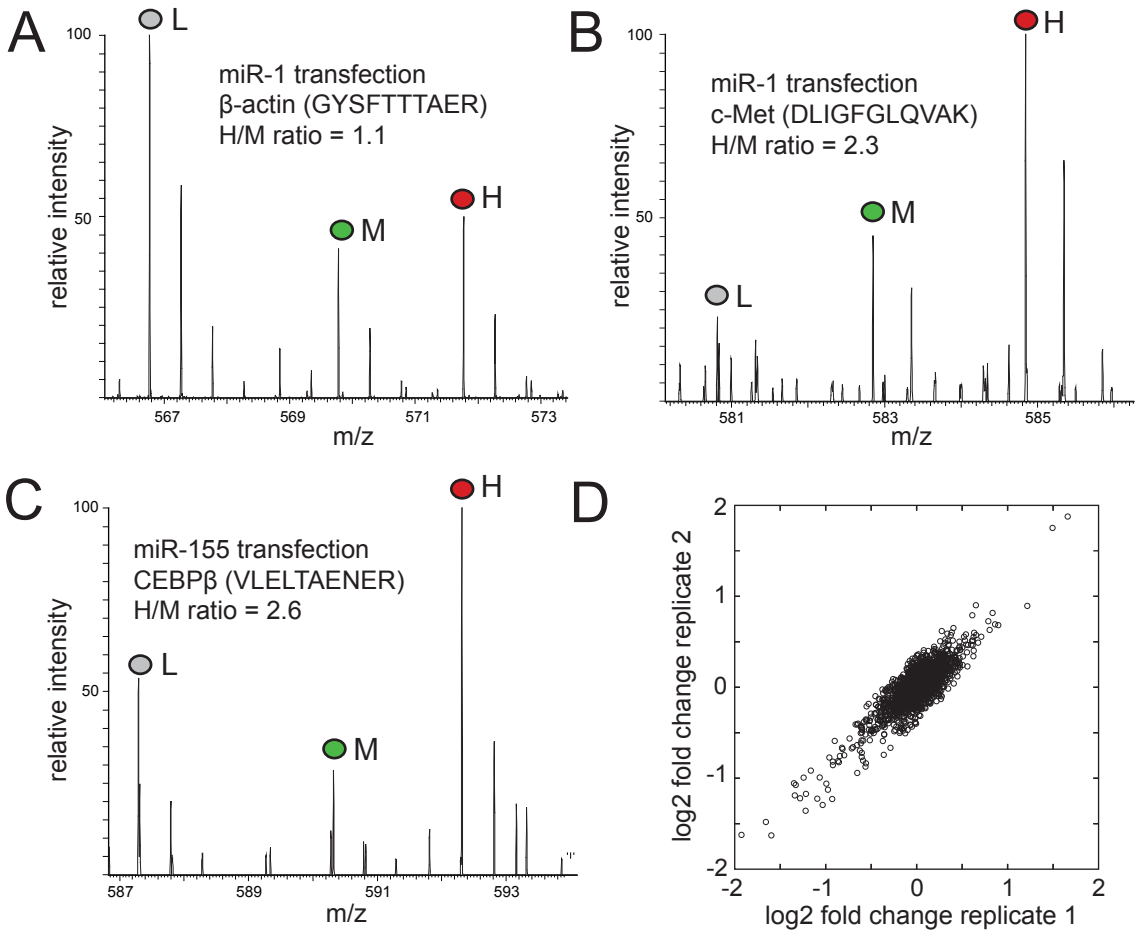


Fig. III.8 | Exemplary mass spectra of identified peptides (sequences in parentheses). (A) Production of most proteins remained unaltered upon miRNA transfection as shown for a β-actin-derived peptide. (B, C) In contrast, synthesis of c-MET, EGFR and CEBP-β was reduced by miR-1 or miR-155 over-expression. (D) Analysis of biological replicates of miR-1 transfections measured by pSILAC revealed high reproducibility of pSILAC.

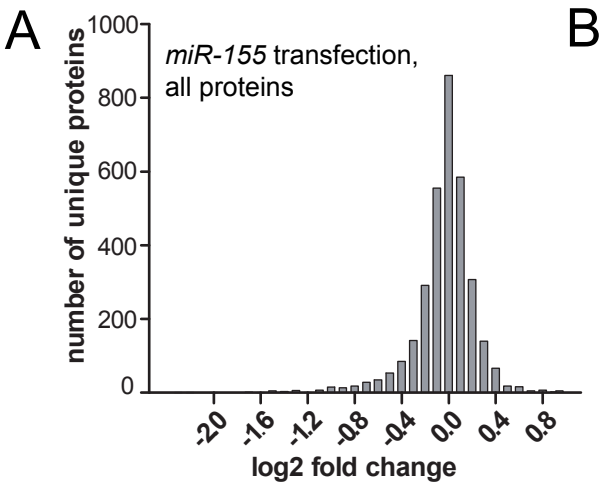
III.2.2. Sequence characteristics associated with reduced protein synthesis

Remarkably, pSILAC measurements revealed that basically all individually over-expressed miRNAs had rather mild effects on the proteome, similar to what has already been observed for the transcriptome (Lim *et al.*, 2003). Exemplary for miRNA-155, Fig. III.9 A shows that overall changes in protein synthesis of ~ 3,300 proteins were rarely beyond 2-fold. Yet closer inspection of the histogram indicates that the distribution of changes is asymmetric and biased towards negative fold changes. Thus, a single miRNA obviously repressed either directly or indirectly the production of several hundred proteins.

Since miRNAs are thought to regulate target mRNAs by binding to *cis*-regulatory sites in 3'UTRs, a linear regression-based computational analysis was conducted to identify putative 3'UTR sequence motifs (cf. Material & Methods II.10) that correlated with the observed overall changes in protein production. More precisely, the applied algorithm termed REDUCE (Bussemaker *et al.,* 2001) performed an unbiased screen for all possible RNA nucleotide motifs of length one to six, assuming a model in which upstream motifs contribute additively to the log-expression level of a gene.

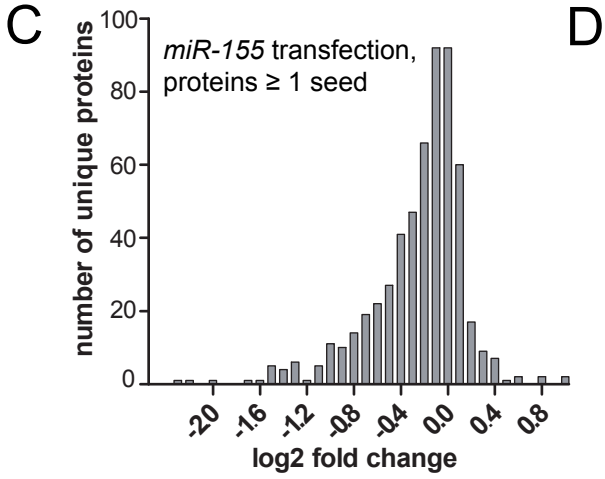
Notably, for each of the five miRNAs, the most significant motif of all possible 5,460 motifs was precisely the seed of the respective miRNA (cf. Fig. III.9 B). For instance, the top ranking motif for the miRNA-1 over-expression experiment given by "CAUUCC" was found in 714 3´UTR sequences and was well correlated (Regression coefficient of -0.29) with down-regulation of protein production measured by pSILAC. Exemplary for miRNA-155 over-expression, plotting the histogram of all fold changes for proteins that contain at least one seed in their mRNA 3' UTRs clearly revealed the correlation between seed enrichment and repression in protein production (cf. Fig. III.9 C). Analysis of seed density, that is, the number of seeds per nucleotide, revealed that the presence of the seed in 3´UTR exerted the strongest negative effect on protein synthesis (cf. Fig. III.9 D). In contrast, the same motif analysis in 5' UTRs had no significant results, and searching in coding sequences (CDS) yielded the seed in only two experiments (let-7b, miR-16).

Towards a more quantitative description of miRNA mediated regulation of seed carrying mRNAs, the number of proteins with at least one seed site in the 3' UTR of respective mRNAs



B

transfected miRNA	top ranking motif	regression coefficient	p-value	number of 3'UTRs with this motif
miR-1	cauucc	-0.29	$<10^{-30}$	714
miR-16	gcugcu	-0.36	$<10^{-30}$	920
miR-30a	guuuac	-0.16	$<10^{-30}$	580
miR-155	gcauua	-0.39	$<10^{-30}$	542
let-7b	uaccuc	-0.20	$<10^{-30}$	480



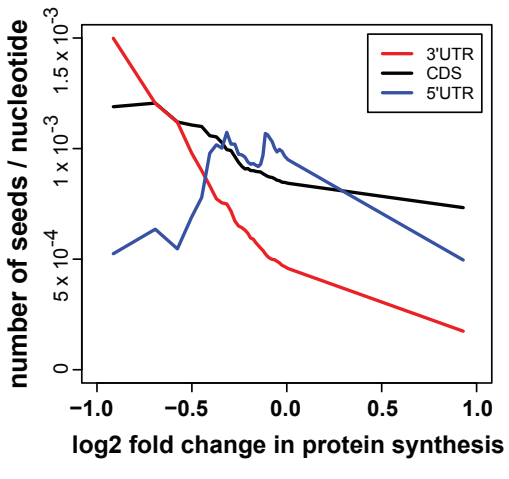


Fig. III.9 | miRNAs down-regulated production of hundreds of proteins. (A) Histogram of changes in protein synthesis of 3,299 proteins in HeLa cells upon miR-155 over-expression. (B) An unbiased search for 3' UTR motifs that correlate with pSILAC fold-changes yielded precisely the miRNA seed sequences. (C) Histogram of proteins with at least one miR-155 seed in corresponding 3´UTRs displayed a clear tendency towards down-regulation by miR-155 over-expression. (D) Analysis of seed frequencies in coding sequences (CDSs), 5' untranslated regions (UTRs) and 3' UTRs of mRNAs and changes in protein synthesis upon miR-16 over-expression. (*Fig. III.9 B and D were contributed by Zhou Fang, bioinformatician at the Berlin Institute for Medical Systems Biology, BIMSB, Robert-Rössle Str. 10, D-13125 Berlin, Germany*)

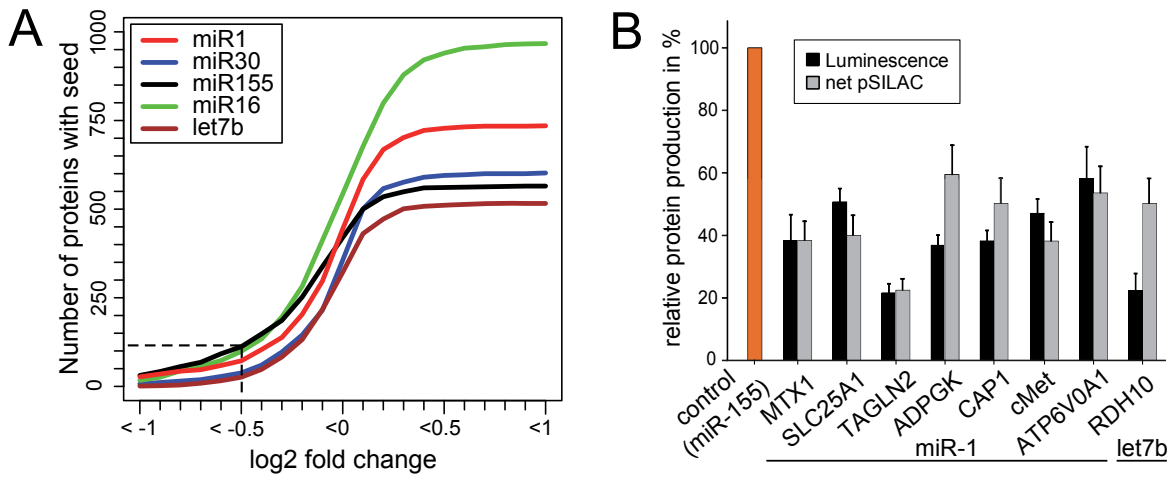


Fig. III.10 | The miRNA seed explained a large fraction of down-regulated protein synthesis. (A) Cumulative number of proteins with seeds as a function of changes in their production. For a given cut-off, this indicates the number of down-regulated seed containing proteins (shown for -0.5 log2 fold-change). (B) Dual luciferase reporter assays for 3' UTR mediated regulation by miRNAs (± s.d., n=3). "Net pSILAC" refers to the difference of pSILAC log2 fold-changes for the miRNA (miR-1 or let7b) and the control (miR-155) (error bars show 95% c.i.; cf. Material & Methods II.13). *(Fig. III.10 A was contributed by Zhou Fang, bioinformatician at the Berlin Institute for Medical Systems Biology, BIMSB, Robert-Rössle Str. 10, D-13125 Berlin, Germany*)

was plotted as a function of their pSILAC fold changes (cf. Fig. III.10 A), separately for each miRNA. Taking all five tested microRNAs together, the production of more than 300 proteins with seeds was repressed by at least 30% (log2-fold change < -0.5). Compared to all proteins down-regulated by at least this much, that is, irrespective of the presence or absence of the seed, the roughly 300 seed-carrying proteins amounted to roughly 60-70%. Hence, it remains an open question how many proteins without a seed are yet direct targets. Nevertheless, the pSILAC approach clearly generated lists of proteins enriched in direct targets, i.e. targets were the presence of a seed mediated target recognition.

In order to validate that the presence of seed sites in 3´UTRs correlated with protein down-regulation measured by pSILAC, dual luciferase reporter constructs were generated (cf. Material & Methods II.11). Eight arbitrarily selected, endogenous 3´UTRs with seeds for either miRNA-1 or let-7b were individually cloned in reporter plasmids downstream of a luciferase gene to analyze the seed-mediated impact of microRNAs on translation. Luciferase reporters were co-transfected with either miRNA-1 or let-7b in HeLa cells, while in control cells miRNA-155 was used instead (cf. Material & Methods II.12 and II.13).

Given the fact that the detection of changes in protein production works fundamentally different in dual luciferase assays, the results correlated surprisingly well with measured pSILAC values (cf. Fig. III.10 B). Of note, only for two genes (ADPGK and RDH10) pSILAC measurements deviated about 20% from luciferase measurements, thereby further confirming the validity of the applied approach.

III.2.3. miRNAs translationally repress many direct target genes

Since the pSILAC approach detects changes at the final step in expression of protein-coding genes, transcriptional as well as post-transcriptional regulation is captured. With regard to

the latter one, detection is irrespective of whether regulation occurs by changes in mRNA stability or at various steps during translation (e.g. initiation or elongation). In order to decipher how much control microRNAs exert explicitly on the transcriptional and translational level, information on transcript levels is needed. To this end, in all pSILAC experiments the mRNA fold-changes between miRNA transfected and control samples were recorded in parallel by Affymetrix microarrays (cf. Material & Methods II.8 and II.15) at the beginning of the pulse labeling (t_1 = 8 h) and at the end (t_2 = 32 h).

Exemplary for miR-1, Fig. III.11 A and B depicts the relation between miRNA induced fold-changes in protein production (measured by pSILAC) and at mRNA levels (measured with microarray) for thousands of genes in scatter plots, separately for t_1 and t_2. Irrespective of the time point, only very few genes had fold-changes of unequal sign and reasonable magnitude (≥ 1.3 fold), indicating that changes in protein production were also reflected to a reasonable

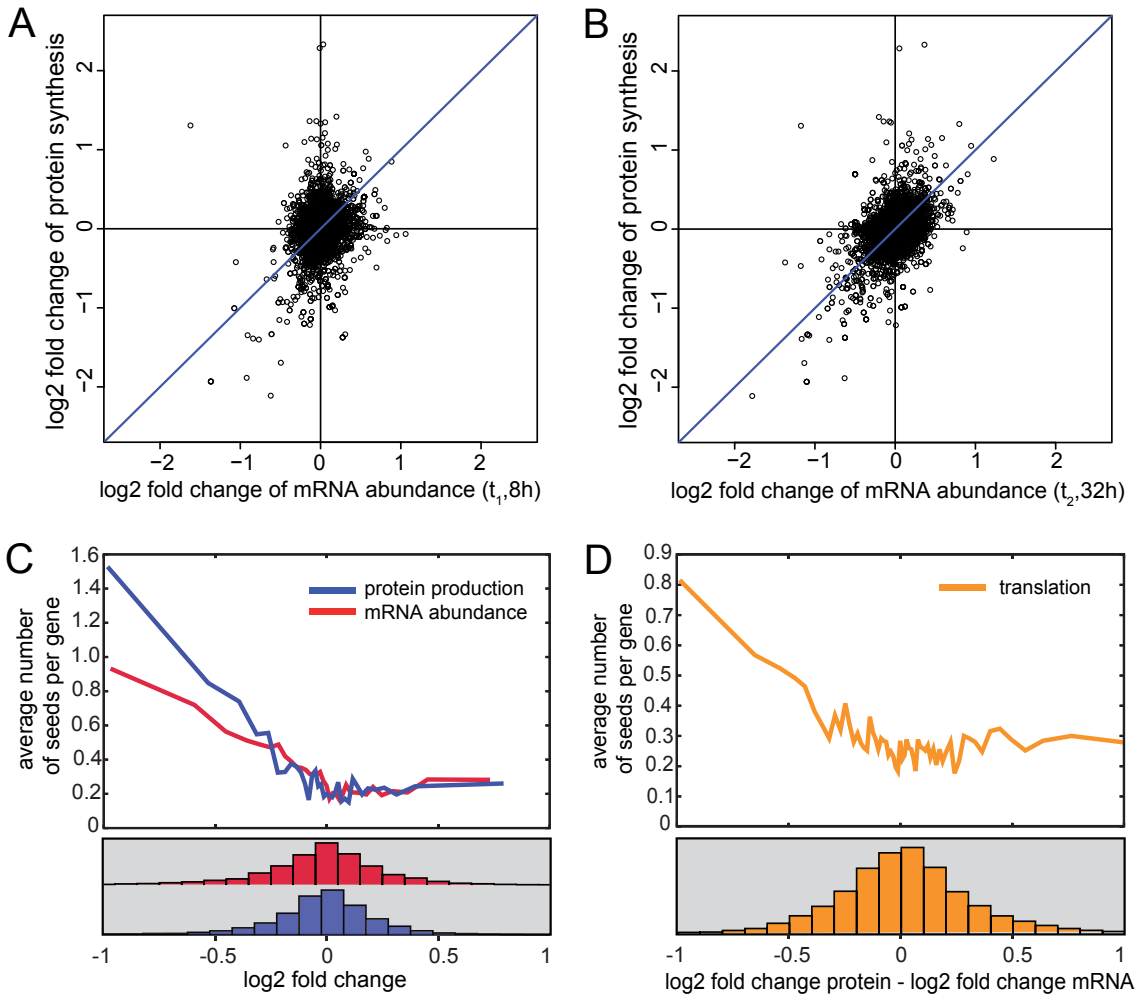


Fig. III.11 | miRNAs inhibit translation on a genome-wide scale. (A) Changes in protein production between 8 and 32 h after miR-1 transfection with mRNA fold-changes at 8 h reveal poor overall correlation. (B) mRNA levels at 32 h correlate remarkably well with changes in protein synthesis. (C) Overall fold-changes of mRNA levels and protein synthesis are similar (histograms). Reduced protein production and mRNA levels correlate with seed frequency (curves represent proteins ranked by fold-change and grouped into bins of 250) (D) Translational repression by miRNAs is revealed by subtracting mRNA log changes from log changes in protein production. Increased seed frequency, averaged as in (C), correlates with translational repression. Results shown for pooled data (C, D) after discarding genes with mRNA and pSILAC changes of unequal sign. *(Fig. III.11 C and D were contributed by Nikolaus Rajewsky, scientific head of Berlin Institute for Medical Systems Biology, BIMSB, Robert-Rössle Str. 10, D-13125 Berlin, Germany*)

extent at steady-state mRNA levels. This applied in particular to the later time point t_2 were the correlation between mRNA fold-changes and pSILAC fold-changes became better compared to t_1. Several genes with down-regulated protein production but little mRNA fold-changes at t_1 obviously shifted towards greater mRNA fold-changes at t_2. Hence, at the first time point t_1 miR-1 had a stronger repressive effect on protein production than on mRNA levels. Similar overall effects could be obtained for other miRNAs (cf. Supplementary Fig. VII.2). Altogether, the considerable scatter given by changes in protein production and mRNA levels indicated substantial and widespread post-transcriptional regulation of gene expression upon microRNA over-expression.

As already apparent from the scatter plot at t_2, closer inspection of fold changes measured by either microarray or pSILAC revealed similar overall frequency distributions (cf. Fig. III.11 C, lower panel). Plotting the average number of seeds per gene in bins of 250 genes against their changes in mRNA levels and protein production revealed that seed frequency was higher for more highly down-regulated genes (cf. Fig. III. 11 C). More precisely, for down-regulated genes log2 fold-changes in protein production and mRNA abundance were linearly correlated with the average number of seeds present in corresponding 3´UTRs. This confirms that the presence of seed sites is crucial for microRNA mediated down-regulation of gene expression. Unexpectedly, the slope of the curve representing pSILAC fold changes in Fig. III.11 C (upper panel) was steeper than for corresponding changes in mRNA levels. This finding suggests that on a genome-wide level, miRNAs apparently exert a stronger direct effect on the down-regulation of protein production than on mRNA levels alone. To quantify this effect more directly, for each gene the log2 fold change in mRNA level were subtracted from the log2 fold change in protein synthesis and plotted as a function of average seed numbers per gene (cf. Fig. III.11 D). In doing so, the linear decline of the obtained curve towards the regime of equal fold-changes revealed for the first time ever on a genome-wide scale the degree of translational control exerted by a single microRNA (miRNA-1). Hence, in addition to mediating mRNA down-regulation presumably via mRNA destabilization (Lim *et al.,* 2005), the seed also mediates direct repression of protein translation for hundreds of genes.

III.2.4. An endogenous miRNA knock-down confirms overexpression experiments

Transfection of single microRNAs into HeLa cells had a widespread impact on protein synthesis. However, one could argue that ectopic over-expression of large amounts of a miRNAs represents a non-physiological condition thereby challenging the findings. To address this issue, the opposite strategy, that is, the knock-down of a highly expressed endogenous miRNA was performed (cf. Material & Methods II.8). To this end, locked nucleic acids (LNAs) (Stenvang *et al.,* 2008; Wahlestedt *et al.,* 2000) were used to knock-down miRNA let-7b followed by measuring changes in protein production by pSILAC.

When comparing let-7b over-expression and knock-down a remarkable anti-correlation for changes in protein production was apparent (cf. Fig. III.12 A). Anti-correlation was even more

evident when averaging fold-changes for the over-expression and knock-down experiment over bins of 20 genes (cf. Fig. III.12 A insert). Though overall log2 fold changes upon knock-down were somewhat smaller as in the over-expression experiment, flipping the sign of pSILAC fold changes revealed converse regulation for the vast majority of proteins. These findings strongly suggest that the performed over-expression experiments of all tested microRNAs did also provide physiologically relevant results.

Intriguingly, anti-correlation did not only apply to seed-containing genes but also to the majority of all ~2,700 proteins quantified in both experiments. For example, when considering all ~130 proteins with a fold-change of at least 15% in both the over-expression and knock-down experiment, the vast majority of them was “consistently” regulated, that is, those genes were up in one of the experiments but down in the other irrespective of seeds (cf. Fig. III.12 B). Hence, both direct and indirect effects were obviously reversed.

Altogether, the results indicated that up- and down-regulation of stationary let-7b levels had largely complementary effects on the proteome, i.e. let-7b levels can obviously tune protein production from thousands of genes.

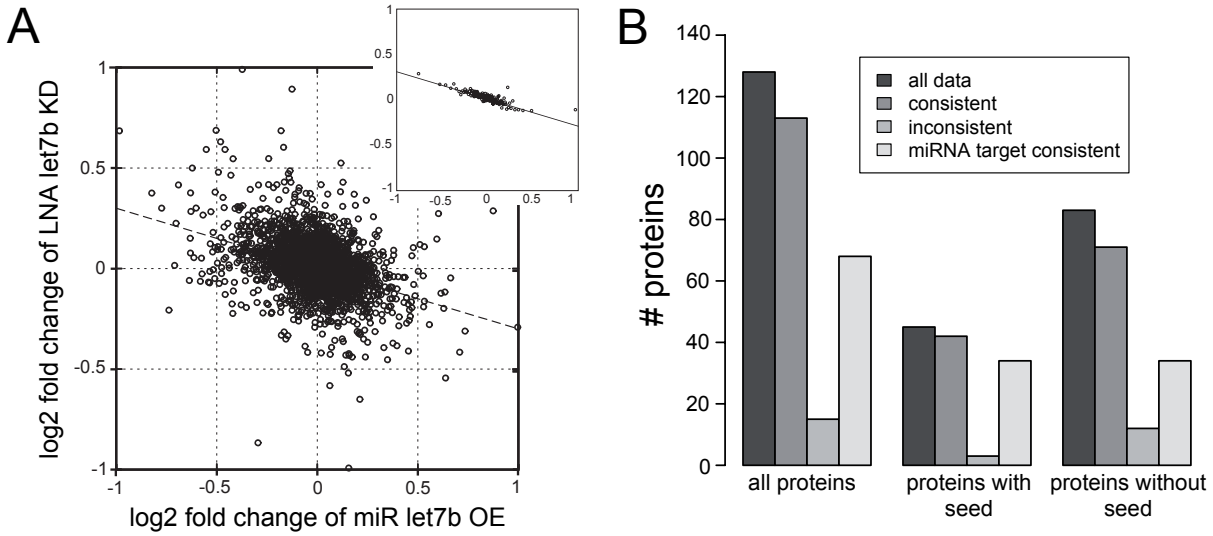


Fig. III.12 | Endogenous knock-down of miRNA-let 7b. (A) Scatter plot of changes in protein production in the let-7b over-expression (OE) versus the let-7b knock-down (KD) experiment. The insert shows the same data, averaged over bins of 20 genes. (B) “consistent” refers to proteins with pSILAC fold-changes (of at least 15%) that were up-regulated in one experiments but down in the other, “inconsistent” to all other cases. “miRNA target consistent” is the subset of “consistent” proteins which were down-regulated in the over-expression experiment but up in the knock-down.

III.3. Genome-wide parallel quantification of mRNA and protein turnover and levels in mammalian cells

III.3.1. Combined metabolic labeling of newly-synthesized RNA and proteins enables determination of half-lives

In order to measure protein and mRNA turnover in the exact same population of exponentially growing, non-synchronized mouse fibroblasts (NIH3T3 cells), metabolic pulse labeling with heavy-stable amino acids and 4sU were simultaneously applied (cf Fig. III.13 and Material & Methods II.17).

III.3.1.1 Determination of cellular protein half-lives

Subconfluent cells grown in light (L) SILAC medium were transferred to heavy (H) SILAC medium and incubated for different lengths of time to account for heterogenous protein turnover. While low turnover proteins require labeling times of several hours to arrive at protein amounts in the heavy form that can be accurately quantified by mass spectrometry, the opposite applies to proteins with relatively short half-lives. For high turnover proteins, short-term labeling avoids that the heavy protein form vastly exceeds the light form and hence limit the traceability of the

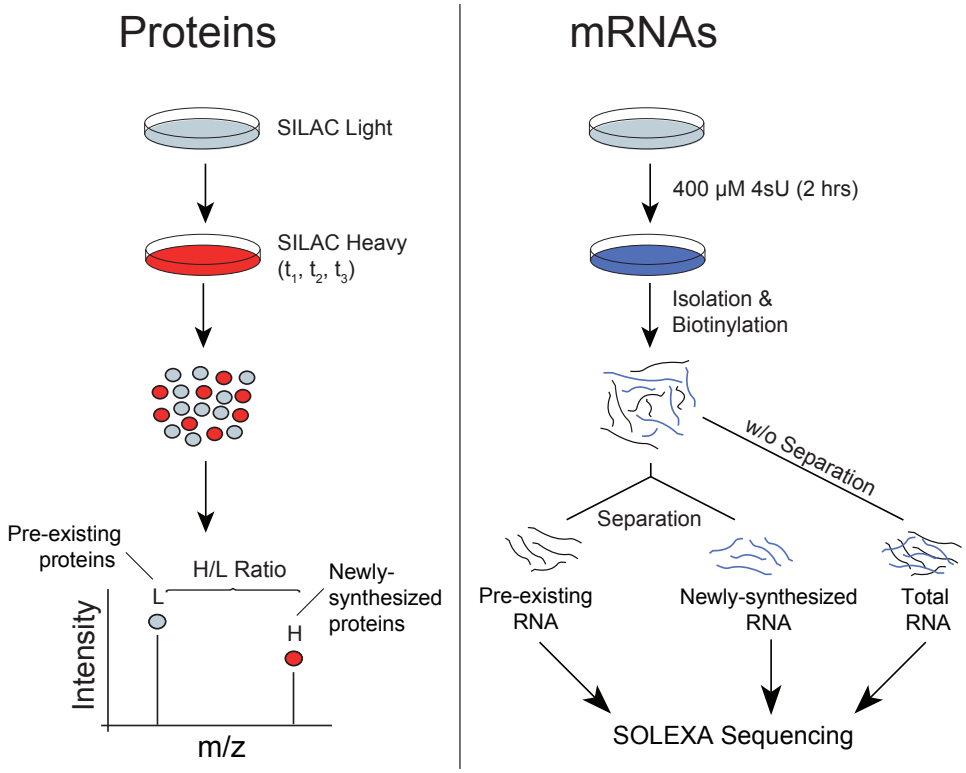


Fig. III.13 | Parallel quantification of mRNA and protein turnover. Mouse fibroblasts were simultaneously pulse-labeled with heavy-stable amino acids (SILAC, left) for different lengths of time (t_1=1.5 h, t_2=4.5 h and t_3=13.5 h) and the nucleoside analogue 4-thiouridine (4sU, right) for 2 h. Protein turnover was quantified by mass spectrometry based on H/L ratios reflecting individual protein synthesis. For determination of transcript half-lives, total RNA was fractionated into newly-synthesized and pre-existing RNA which were separately subjected to next-generation sequencing (SOLEXA Sequencing).

H/L ratio. To cover a reasonable time range, protein samples were harvested at three different time points (1.5 h, 4.5 h and 13.5 h) upon transfer to heavy SILAC medium and analyzed by high resolution mass spectrometry (cf. Material & Methods II.24.1 and II.24.4).

Representative mass spectra of peptides derived from two proteins with distinct turnover rates indicated by deviant incorporation rates of the heavy label are depicted in Fig. III.14 A and B. The heavy form of the Ribonucleoside-diphosphate reductase subunit M (Rrm2) was rapidly increasing over time compared to the light peptide, resulting in a H/L ratio of greater then ten after labeling for 13.5 h. As opposed to this, the Histone H1.2 (Hist1h1c) displayed a rather moderate shift from the light to the heavy form, denoting the protein´s turnover to be lower than that of Rrm2.

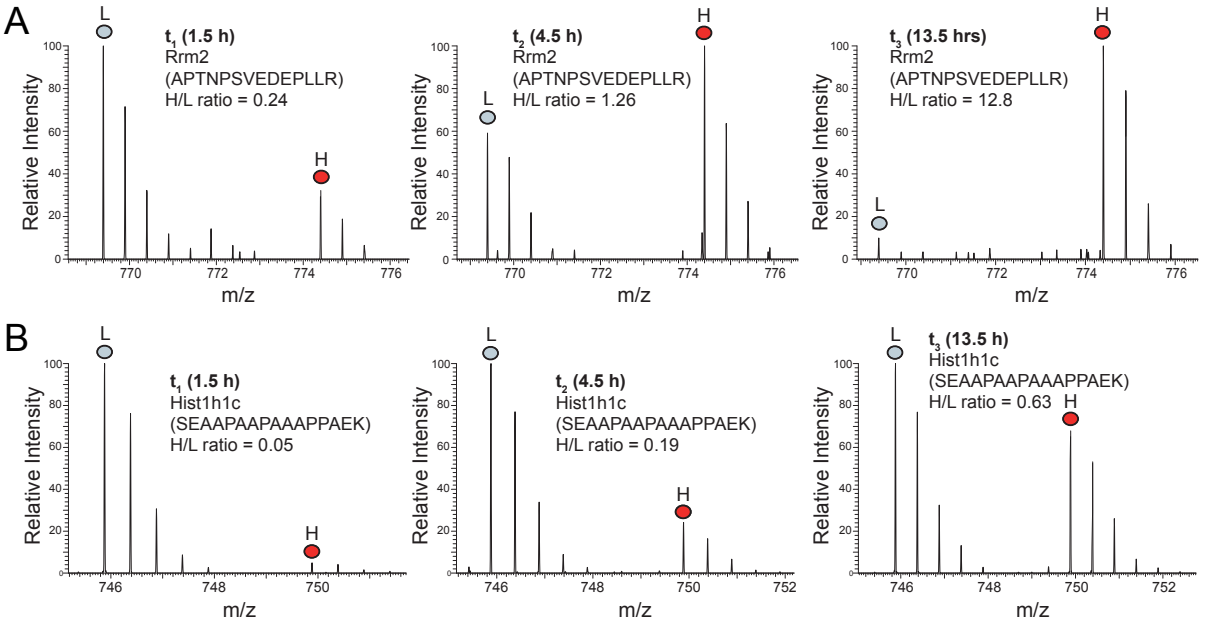


Fig. III.14 | Mass spectra of peptides derived from proteins with different turnover times. While the H/L ratio of the Ribonucleoside-diphosphate reductase subunit M (Rrm2) showed a rapid increase over time (A), the Histone H1.2 (Hist1h1c) exhibited a rather moderate augmentation (B). The shift from the light to the heavy peptide form can therefore be used as a measure for protein turnover.

Robust quantification, however, relies not only on the instrument´s capability to precisely measure H/L ratios, it also necessitates that the ratios themselves are reliable. In this regard, particularly the height of the light peptide peak can be over-estimated for two reasons: First, cellular degradation of light proteins after transfer to heavy medium can result to some degree in recycling of light amino acids. Thus, newly-synthesized proteins will not exclusively appear in the heavy form. Second, remaining light amino acids from the cellular precursor pool can also be used for protein synthesis, again resulting in an under-estimation of the heavy peptide form. Since both events potentially compromise subsequent calculation of protein half-lives, they had to be thoroughly addressed.

Fortunately, the degree of light amino acids (re-)used for *de novo* protein synthesis in cells transferred to heavy SILAC medium can be evaluated by closer inspection of mass spectra obtained from heavy, missed cleaved tryptic peptides (trypsin cleaves with high specificity

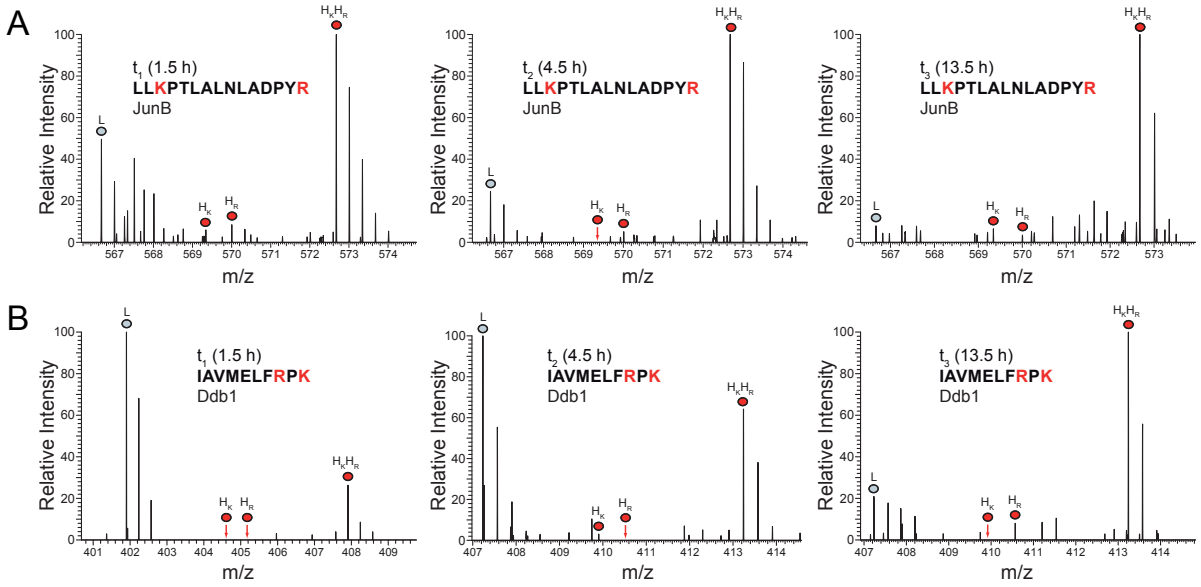


Fig. III.15 | Uniform incorporation of heavy amino acids. Protein turnover can be under-estimated due to the presence of light (L) amino acids that remain in the precursor pool after transferring cells to heavy (H) SILAC medium (e.g. by recycling). The degree of L amino acids incorporated after transferring cells to H medium can be evaluated by inspection of mass spectra of missed cleaved tryptic peptides which contain two labeled amino acids (K=lysine, R=arginine). In case of complete labeling newly-synthesized peptides should contain only heavy lysine and heavy arginine residues (H_KH_R). Additional peaks containing only one heavy amino acid (H_K or H_R, marked in red) are indicative of light amino acids incorporated after medium exchange. The figure shows mass spectra of missed-cleaved peptides from a high (JunB, A) and a medium (Ddb1, B) turnover protein at all three time points. In some cases, partially heavy labeled peptides were observed.

after lysine (K) and arginine (R) residues). Under ideal labeling conditions, such missed cleaved peptides carry one additional heavy lysine or heavy arginine moiety (H_KH_R) which doubles the mass difference to the corresponding unlabeled, light peptide (cf. Fig. III.15 A and B). However, if only partial heavy labeling of a missed cleaved peptide due to the presence of light amino acids was achieved, an additional peptide mass between the doubly light labeled and the doubly heavy labeled peptide occurs (H_K or H_R). The extent of such a missed cleaved and partially heavy labeled peptide is indicative of the magnitude of (re-)use of light amino acids for protein synthesis. Exemplary for the transcription factor jun-B (JunB) and the DNA damage-binding protein 1 (DdB1), Fig. III.15 A and B show that in fact partially heavy labeled peptides could be observed. Yet these peaks were small, i.e. less than 10% of the doubly heavy peaks, suggesting that only a small fraction of newly synthesized peptides contained recycled and/or remaining light precursors.

In total, 1,471,375 fragment spectra were acquired that resulted in 229,985 peptide identifications (84,924 unique peptide sequences). These peptides were assigned to 6,445 unique proteins (FDR < 1%, cf. Material & Methods II.24.5). At least three H/L peptide ratios per protein over all time points were required for quantification, which was met by 5,279 proteins. Since most of the proteins were quantified at two or all three time points, protein half-lives were calculated based on linear regression applied to all H/L ratios obtained for each protein (cf. Material & Methods II.29.1). Fig. III.16 A exemplifies linear regression for Rrm2 and Hist1h1c. Though for most proteins the correlation was very good, only proteins with an excellent correlation ($R^2 \geq 0.9$) were kept for further analysis (cf. Fig. III.16 B). By doing so, 5,066 proteins remained and were used for individual half-life calculations. Of note, half-

lives were also computed for proteins quantified at a single time-point only. Although linear regression was not applicable in these cases, robustly quantified single H/L protein ratios, that is, requiring at least three peptide ratios at that time point allowed to include them in the analysis (cf. Material & Methods II.29.1). Individual H/L ratios obtained from up to the three time points reflect the net outcome of a protein´s synthesis and degradation. Assuming exponential decay on the one hand and a doubling of the overall protein amount during one cell cycle on the other hand enabled the calculation of individual degradation rate constants (k_d) and eventually protein half-lives (cf. Material & Methods II.29.1).

The overall distribution of cellular protein half-lives is depicted by the histogram in Fig. III.16 C. Protein half-lives ranged from below one to several hundred hours with a median of 45 h. For validation purpose, protein stability of proteins spanning a range from one to 60 h was assessed by cycloheximide-chase experiments (cf. II.16 and II.25 and) (cf. Fig. III.16 D). Estimated protein stabilities of selected candidates were overall in good concordance with turnover times obtained by the SILAC approach. For instance, the cycloheximide-chase experiment denoted

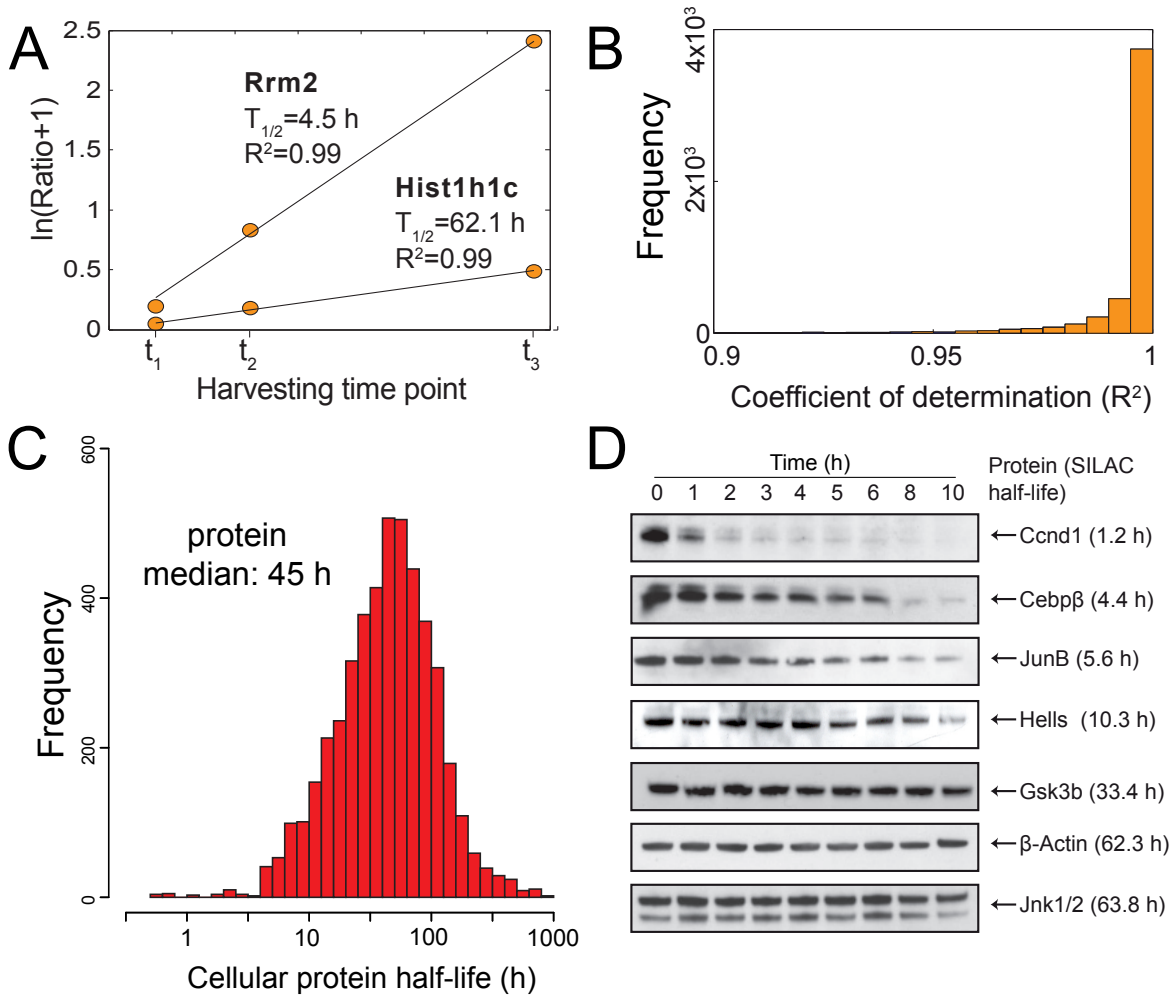


Fig. III.16 | Calculation and validation of protein half-lives. (A) Protein half-lives were calculated from log H/L ratios at all three time points using linear regression as shown for the proteins Rrm2 and Hist1h1c. As a quality test of linear regression, the coefficient of determination (R^2) was calculated. (B) Protein half-lives were only calculated for fits that satisfied $R^2 \geq 0.9$. This criteria was met by 4,906 proteins. (C) Histogram of all calculated protein half-lives. (D) Comparison of protein half-lives measured by SILAC and traditional cycloheximide-chase experiments. *(Fig. III.16 A and B were contributed by Dorothea Busse, scientist in the group of Jana Wolf termed "Mathematical Modelling of Cellular Processes" at the Max-Delbrück-Centrum for Molecular Medicine (MDC, Robert-Rössle Str. 10, D-13125 Berlin, Germany)*

the Cyclin D1 protein (Ccnd1) to be highly unstable, i.e. the protein was almost fully degraded after 2 h. Similarly, the CCAAT/enhancer-binding protein beta (Cebpβ) was barely detectable 8 h upon blockage of protein synthesis via cylcloheximide. Both observations matched very well to the SILAC-based determination of half-lives, where for Cyclin D1 and Cebpβ turnover times of 0.8 h and 6 h were computed, respectively.

III.3.1.2 Determination of cellular mRNA half-lives

In parallel to metabollic labeling of proteins, newly synthesized RNA was pulse-labeled for 2h with 4sU essentially as described previously (Doelken *et al.,* 2008) (cf. Material & Methods II.17). Total RNA samples were fractionated into the newly-synthesized and pre-existing RNA fractions (cf. Material & Methods II.18.1 and II.18.2). In order to validate the separation

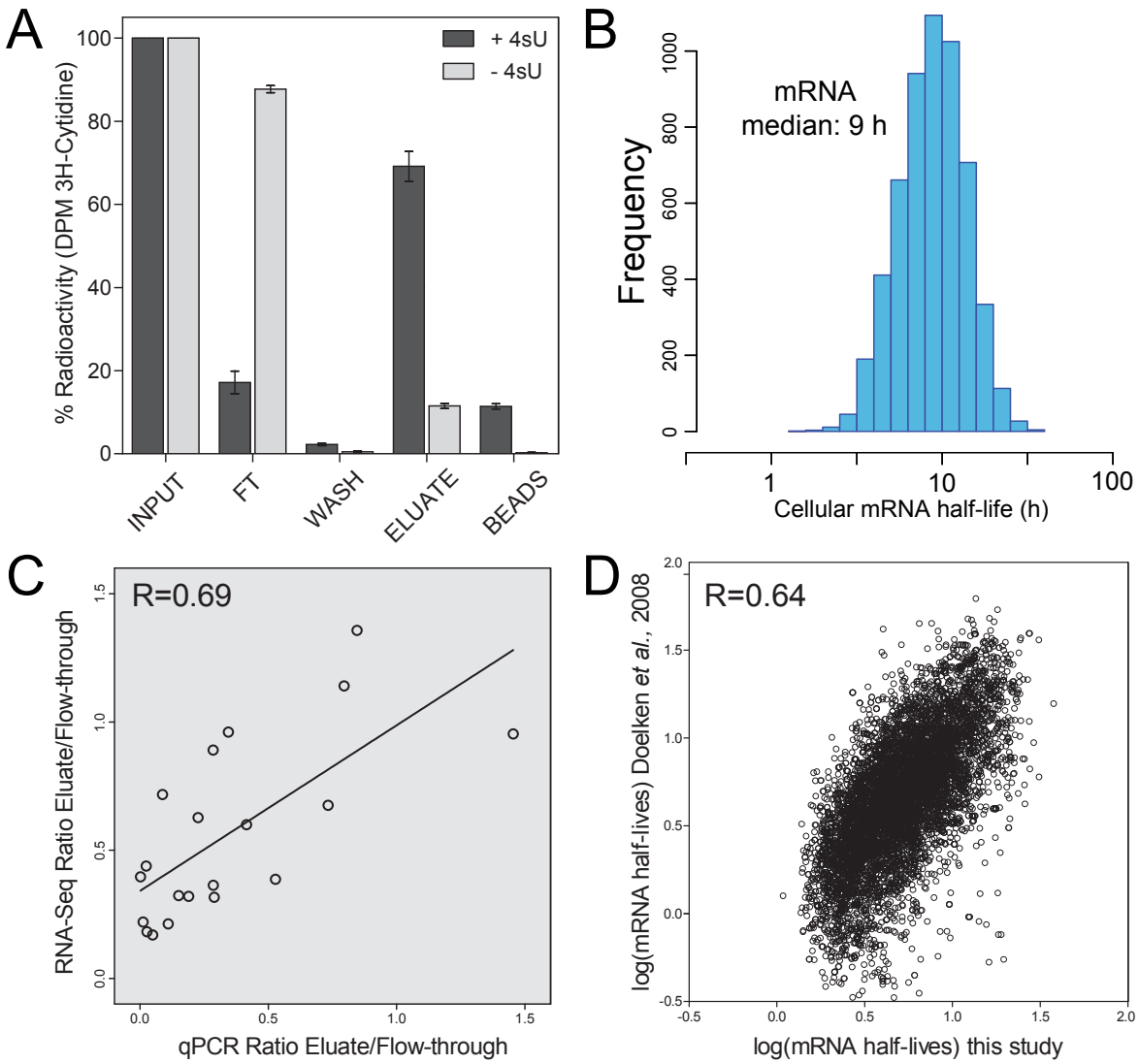


Fig. III.17 | Overall distribution and validation of mRNA half-lives. (A) Separation efficiency of total RNA into newly-synthesized (eluate) and pre-existing (flow through, FT) RNA fractions was validated by combined 4sU and 3H-cytidine labeling. The input refers to the summed radioactivity of all fractions (flow through, wash, eluate and beads) and was set to 100%. (DPM = decay per minute) (B) Overall distribution of measured mRNA half-lives with a median half-life of 9 h. (C) A total of 19 RNA-Seq measurements of eluate/flow-through ratios (i.e. newly-synthesized RNA/pre-existing RNA ratios) were reexamined using quantitative RT-PCR (qRT-PCR). (D) Comparison of mRNA half-lives in NIH3T3 cells obtained in this study to publicly available half-lives measured with microarrays (Doelken *et al.*, 2008) for the same cell line.

efficiency newly-synthesized RNA was co-labeled by 4sU and radioactive cytidine (3H-cytidine, cf. Material & Methods II.18.3). After biotinylation and column-based affinity purification up to 70% of the radioactivity was found in the fraction of newly-transcribed RNA (Eluate, cf. Fig. III.17 A). By contrast, the bulk of radioactively labeled RNA was found in the flow through fraction when labeling with 4sU was omitted.

Both fractions (newly-synthesized and pre-existing fractions) and the total RNA sample were analyzed by mRNA sequencing on a Solexa platform (cf. Material & Methods II.19). mRNA half-lives were calculated based on the ratios of newly-synthesized RNA/total RNA and pre-existing RNA/total RNA essentially as described previously (Doelken *et al.,* 2008; cf. Material & Methods II.20). In total, for 12,629 genes mRNA half-lives could be determined. The overall distribution of cellular transcript half-lives for genes also detected at the protein level is shown by the histogram in Fig. III. 17 B. mRNA turnover ranged from 1 to 500 hours with a median of 9 h. For validation purpose, the eluate/flow-through ratios (i.e. the newly-synthesized RNA/pre-existing RNA ratios) measured by RNA-Seq of 19 randomly selected mRNAs were independently quantified by qRT-PCR (cf. Material & Methods II.22). The obtained results agreed well (Pearson R=0.69) with the sequencing data (cf. Fig. III.17 C). In addition, the calculated mRNA half-lives were consistent (Pearson R=0.64) with previously published data for NIH3T3 cells (Doelken *et al.*, 2008) obtained by conventional microarrays (cf. Fig. III. 17 D).

III.3.2. Sequence characteristics affecting mRNA and protein stability

It has been shown that mammalian mRNA and protein stability are regulated, in part, by the presence or absence of regulatory elements (Wilusz *et al.* 2001; Yang *et al.*, 2003; Rogers *et al.,* 1986; Dice, 1987). Such sequence features are thought to be specifically recognized by multiprotein complexes governing cellular RNA or protein degradation (Keene, 2007; Barreau

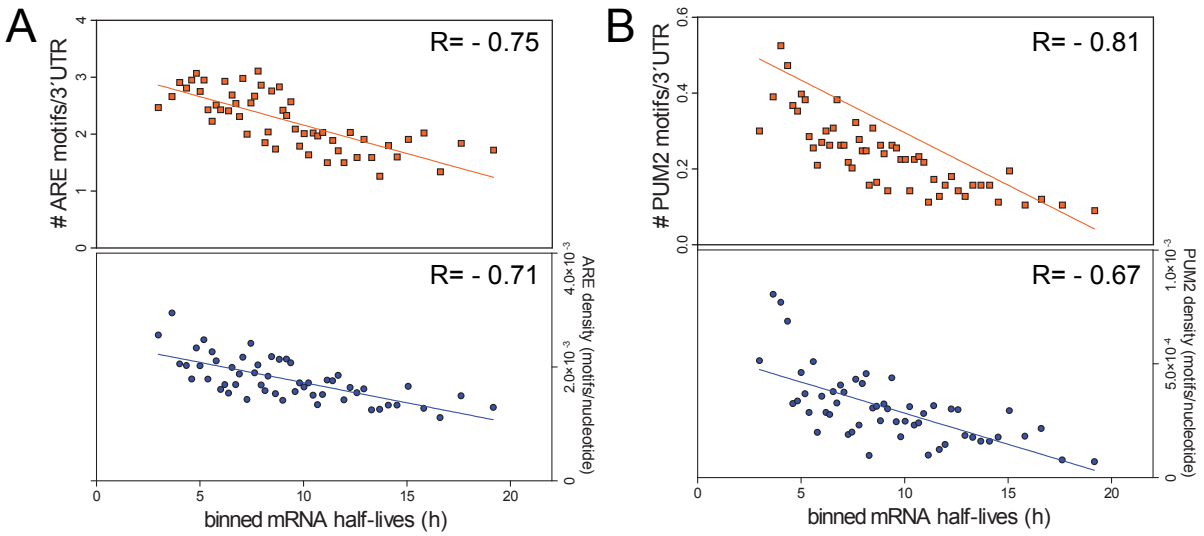


Fig. III.18 | Sequence motifs affecting mRNA stability. mRNA half-lives were binned in groups of 100 genes and plotted against the number of AU-rich elements (AREs) or Pumilio 2 (Pum2) binding sites in corresponding 3´UTR regions (A and B, upper panel). The negative correlation was still clearly apparent after correction for different 3´UTR lengths to account for stochastic motif occurrences (A and B, lower panel).

& Paillard, 2005; Dyson and Wright, 2005).

With regard to mRNA decay, amongst the most intensively studied sequence determinants driving mRNA degradation are *cis*-regulatory motifs termed AU-rich elements (AREs; stretches consisting of mainly adenine and uracil nucleotides) predominantly found in 3´UTR regions. AREs were shown to be bound by numerous proteins to normally shorten mRNA life span (Shaw & Kamen, 1986; Barreau *et al.*, 2006) or, in response to specific extracellular signals also extend mRNA life span (Vasudevan & Steitz, 2007). Similarly, the RNA-binding protein Pumilio 2 (Pum2) has been proposed to control mRNA stability and translation through its sequence-specific interaction with the 3'UTR region of target mRNAs (Wang *et al.* 2002).

In order to test whether the presence of such sequence features also influenced mRNA stability in NIH3T3 cells, individual 3´UTRs were screened for the relatively well-defined AU-rich core element ("ATTTA"; Barreau *et al.,* 2006) and Pumilio2 binding sites ("TGTA(A|T|C|G)ATA" with "I" referring to "or"; Hafner *et al.,* 2010). The number of motifs for each 3´UTR was counted and plotted as a function of binned transcript half-lives (cf. Material & Methods II.31). As shown in Fig. III.18 A and B (upper plots), the number of ARE and Pum2 motifs found in 3´UTRs was negatively correlated with mRNA stability. Of note, the negative correlation also held true when ARE and Pum2 motif densities (i.e. "motifs per nucleotide") were computed to account for stochastic motif occurrences due to vastly different 3´UTR lenghts (cf. Fig. III.18 A and B, lower plots).

Since several sequence features potentially affecting mRNA stability are primarily situated in untranslated regions of mRNAs (Keene, 2007), the lengths of such regions themselves were negatively correlated with mRNA stability. This applied especially to 3´UTR lengths (cf. Fig. III.19 A), and, to a lesser degree also to 5´UTR lengths (cf. Fig. III.19 B). As opposed to this, the lengths of the coding sequence (CDS) only exhibited a positive correlation with mRNA half-lives (cf. Fig. III.19 C).

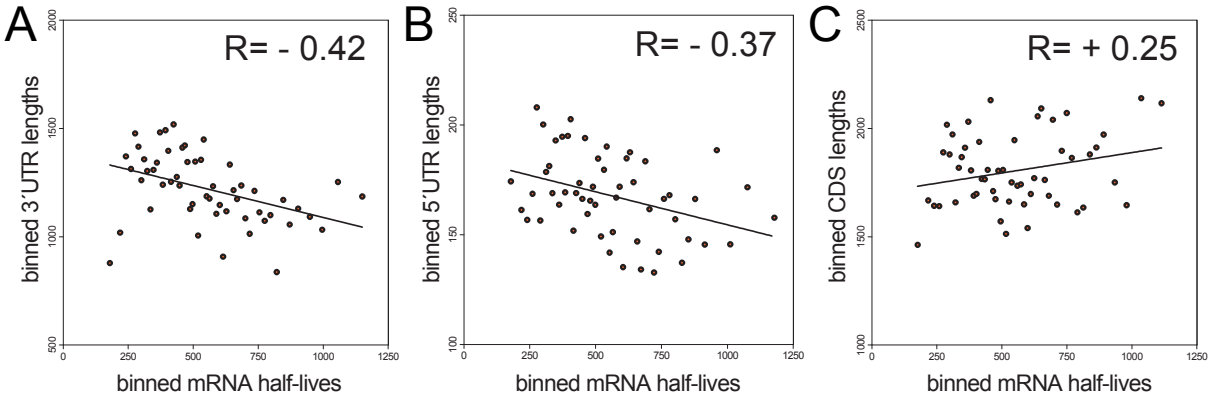


Fig. III.19 | Impact of 3´UTR, 5´UTR and CDS lengths on mRNA stability. Binned mRNA-halflives (in groups of 100 genes) were plotted against corresponding, binned (A) 3´UTR, (B) 5´UTR or (C) CDS sequence lengths.

Amongst different degradation or destruction signals (termed degrons) directing proteins to the ubiquitin-proteasome system, PEST-motifs and intrinsically unstructured regions (IUPs) have been suggested. Proteins containing PEST regions, that is, sequence stretches of 10-50

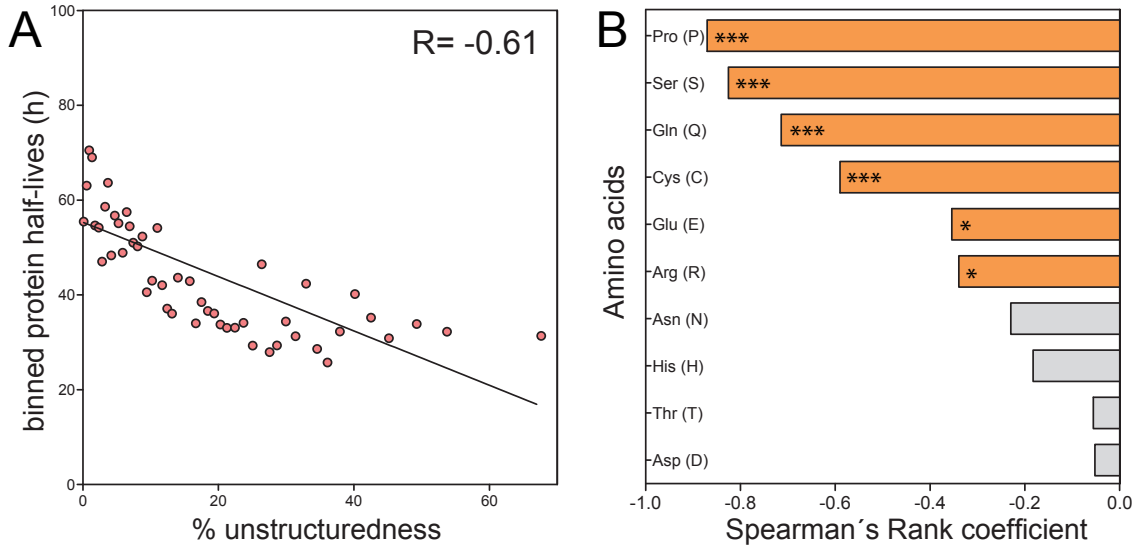


Fig. III.20 |Degradation signals with an impact on protein stability. (A) The degree of disorder of each protein (in bins for 100 genes) was plotted against corresponding, binned protein half-lives. (B) The bar plot shows to what degree relative fractions of amino acids correlate to protein turnover. For instance, the amino acid proline frequently occurs in unstable proteins, given by highly significant Spearman´s Rank coefficient (* indicates p-values < 0.05 but ≥ 0.01; *** indicates p-values < 0.001)

amino acids enriched in proline (P), glutamic acid (E), serine (S), and threonine (T) tend to be turned over rapidly (Rogers *et al.,* 1986; Dice, 1987) although the underlying mechanism remain elusive. Similarly, sequence segments enriched in most polar and charged residues but depleted of hydrophobic residues are considered to correlate with low protein stability. Those sequence stretches are assumed to be largely intrinsically disordered and hence extremely susceptible to proteolytic degradation (Tompa, 2002; Tompa *et al.*, 2008; Gsponer *et al.,* 2008).

The definition of disordered regions of amino acid sequences based on the primary structure alone is, however, a fairly complex endeavour, since also spatial circumstances (i.e. secondary and tertiary structure) considerably impact on protein degradation. As a proxy for a protein´s "unstructuredness", a publicly available mouse protein database was used that predicts potentially disordered residues in a context-specifc manner and also assigns confidence values accordingly (cf. Material & Methods II.31). Comparing the overall unstructuredness of individual proteins with their protein half-life revealed a marked negative correlation. Thus the degree of disorder obviously affects protein stability (cf. Fig. III.20 A). Interestingly, this negative correlation in mouse fibroblasts was stronger than in yeast, where structural disorder served only as a weak signal for intracellular protein degradation (Tompa *et al.*, 2008).

The assignment of PEST motifs in amino acid sequences is also challenging, since the enrichment of respective amino acids cannot be detected by a straightforward search for linear sequence motifs (as for AU-rich elements, see above). Alternatively, relative fractions of amino acids per protein can be computed and compared to protein lifetimes (cf. Material & Methods II.31). The degree of correlation (given by the Spearman´s Rank coefficient) can then be used to investigate whether the presence of certain amino acids, such as P, S, and T in proteins with specific turnover characteristics deviates from mere stochastic occurrence. As shown in Fig.

III.20 B, the relative fractions of amino acids associated with the PEST motif exept for Threonin (T, Thr) were inversely correlated with protein stability indicated by statistically significant Spearman´s Rank coefficients. Hence, due to the high statistical significance of prolin and serine densities, the results support at least partially the destabilizing effect of the PEST motif, which is also in accordance with previous results derived from yeast protein half-lives (Belle *et al.,* 2006).

III.3.3. Estimating cellular protein and mRNA levels

In addition to mRNA and protein half-lives, parallel mRNA sequencing and mass spectrometry also allowed to estimate cellular mRNA and protein copy numbers. Absolute cellular transcript and protein concentrations were deduced from the very same RNA-Seq and mass spectrometry data used for calculation of half-lives. Hence, this is the first time ever that half-lives and levels were quantified for the exact same set of cells, thereby excluding errors oftentimes emerging from integrating experimental data from different sources.

III.3.3.1 Absolute mRNA quantification

Most available mRNA expression data is based on traditional microarrays. Recently, next generation sequencing emerged as more precise and reliable approach for mRNA quantification. The number of reads per kilobase of exon per million mapped sequence reads (RPKM) provides a digital measure of cellular mRNA levels (Mortazavi *et al.,* 2008; Pepke *et al.,* 2009). Combined with information about cellular mRNA content, RPKM values were used to calculate absolute cellular mRNA copy numbers essentially as described previously (Mortazavi *et al.,* 2008; cf. Material & Methods II.21).

III.3.3.2 Absolute protein quantification by intensity-based absolute quantification (iBAQ)

As mentioned in the introduction part, several methods have been developed in order to determine absolute protein levels by mass spectrometry. In essence, spectral counting-based methods like PAI, empAI and APEX rely on the observation that abundant proteins are also represented by a large number of peptides. This results in a higher number of spectral counts in a shotgun experiment compared to proteins with lower concentrations. In contrast, XIC (extracted ion chromatogram)-based approaches make use of a peptide intensities to deduce absolute protein levels.

To systematically benchmark the prevailing methods for absolute protein quantification, a well-defined reference sample was employed. A suitable measure for absolute protein abundance should be independent of protein sequence and provide accurate data over several orders of magnitude. Furthermore, quantification should not be affected by overall sample complexity. Sigma-Aldrich recently developed an improved Universal Proteomics Standard (UPS2) comprising 48 precisely quantified human proteins formulated into a dynamic range

of concentrations spanning six orders of magnitude. These 48 proteins were used either alone or in combination with different amounts of *E. coli* lysate to obtain reference samples of high dynamic range which simulate the complexity of biological probes (cf. Material & Methods II.25 and II.26). All samples were subjected to in-solution digests and analyzed by LC-MS/MS in triplicates (cf. Material & Methods II.24.2 and II.24.4).

Depending on the individual experiment between 20 and 28 of the standard proteins over three to six orders of magnitude were identified with very high confidence (cf. Material & Methods II.24.5). By plotting the known amounts of standard proteins against different measures for absolute protein abundance followed by calculation of the coefficient of determination R^2, the performance of the existing different methods for absolute protein quantification was assessed (cf. Fig. III.21 and Supplementary Fig. VII.3). Since R^2 reflects to what extent the variation in one variable can be explained by changes in another variable, raw spectral counts explained only around 40-60% of the differences in absolute amounts. Correcting spectral counts for the number of observable peptides (PAI) or protein observability scores (APEX) boosted this value to 80-90%. In contrast to these spectral counting-based approaches, the sum of peptide peak intensities explained already 80% of absolute protein abundance levels without any additional correction. Intensities of the three most intense peptides per protein (TOP3) slightly improved the correlation but also resulted in a smaller number of quantifiable proteins as not all proteins were detected with three different peptides (not shown). Intriguingly, the best correlation with absolute protein amounts was achieved by simply dividing the sum of all peptide intensities by the number of theoretically observable peptides. This method was termed iBAQ for intensity-based absolute quantification. Depending on the individual experiment iBAQ explained between

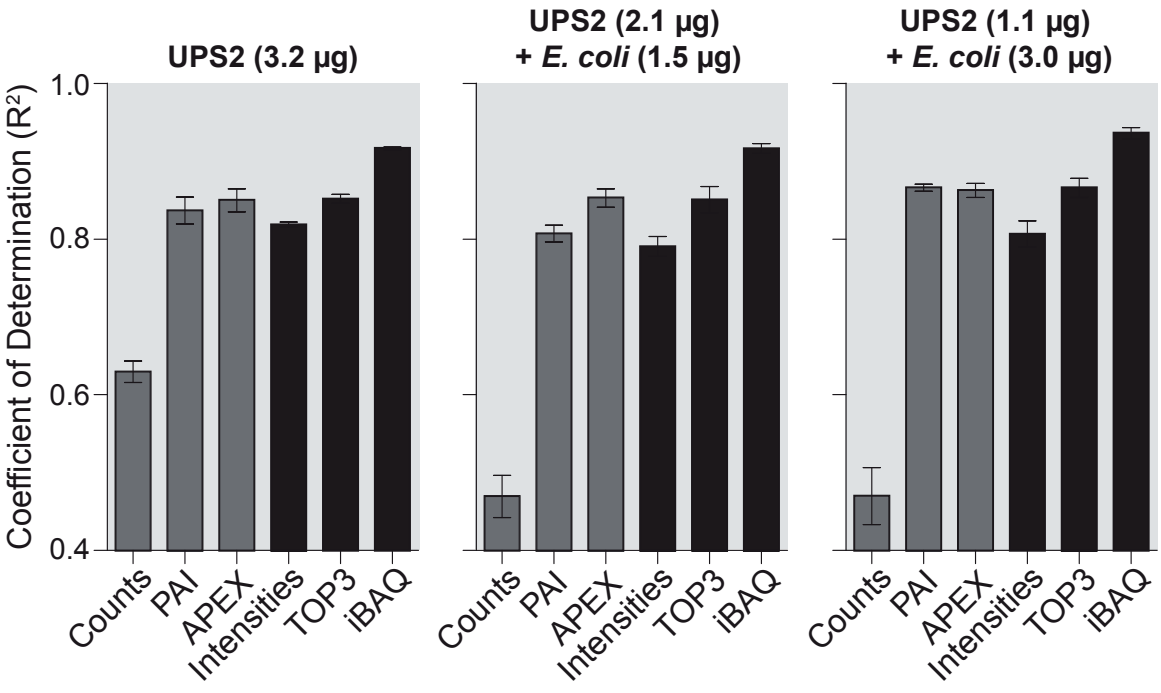


Fig. III.21 | Intensity-based absolute quantification (iBAQ). Correlation between spectral counting-based (grey) or intensity-based (black) methods of absolute quantification and known amounts of standard proteins (error bars: SEM). 48 precisely quantified standard proteins (UPS2) either measured alone or mixed with *E. coli* lysate in different ratios followed by digestion and LC-MS/MS analysis in triplicates.

90 and 95% of the differences in absolute protein amounts. Importantly, iBAQ significantly outperformed the other tested approaches in all experiments ($p < 0.002$, one-tailed Wilcoxon signed rank test using all nine R^2 values from technical and biological replicates).

Since the iBAQ approach was applicable to the mass spectrometry measurements used to compute protein half-lives, an accurate proxy for protein levels could be obtained from the very same data set. By taking into account information about the number of cells used in the experiment, protein concentrations obtained by iBAQ were eventually converted into cellular protein copy numbers (cf. Material & Methods II.26).

III.3.4. Correlation of protein and mRNA levels and half-lives

A direct comparison of mRNA and protein half-lives is depicted in Fig. III.22 A. Transcript lifetimes displayed a significantly smaller spread compared to protein half-lives. While mRNAs spanned a little more than two orders of magnitude, protein turnover was spread over more than five orders of magnitude. Inspection of the histogram showing cellular mRNA and protein copy numbers revealed that proteins with a median copy number per cell of 10,000 were on average ~650 times more abundant than transcripts with a median copy number per cell of only 15 (cf. Fig. III.22 B). Also, overall protein levels displayed a greater dynamic range compared to

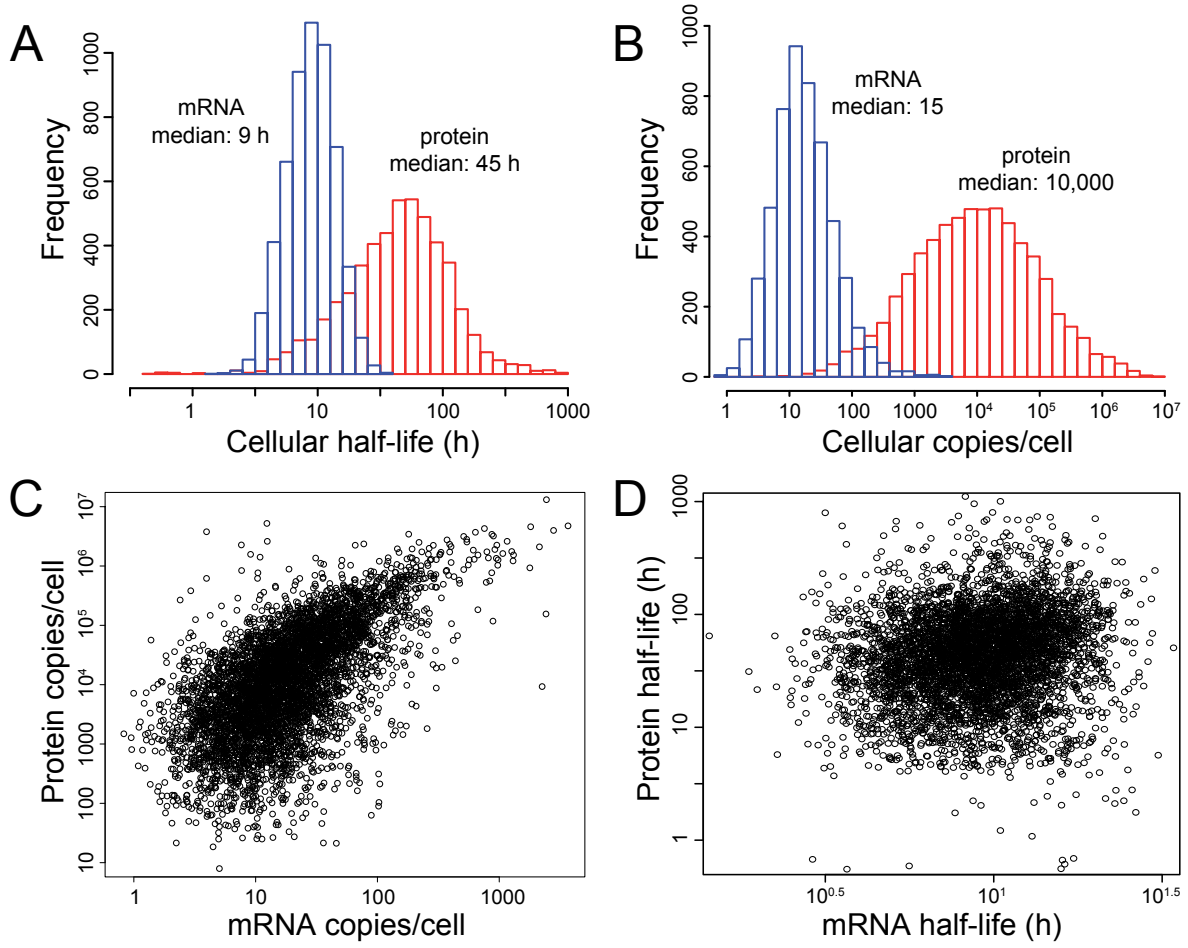


Fig. III.22 | mRNA and protein half-lives and levels. Histograms of mRNA (blue) and protein (red) half-lives (A) and levels (B). Proteins were on average five times more stable and approximately 650 times more abundant than mRNAs and spanned a higher dynamic range. Correlations between mRNA and protein levels (C) and half-lives (D) were analyzed using scatter plots.

overall mRNA levels. While protein concentrations spanned almost six orders of magnitude, mRNA levels were spread over less than four orders of magnitude. Relatively few proteins had less than 100 copies per cell, suggesting that the analysis was somewhat biased against lowly expressed proteins. Comparing mRNA levels of detected and undetected proteins revealed a mild detection bias (cf. Supplementary Fig. VII.4) suggesting that the true dynamic range of protein levels can be expected to be even higher. For that reason, all further analysis were restricted to the subset of genes that was identified at both the mRNA and protein level.

However, apart from the difference in dynamic range, the scatter plot demonstrates that overall mRNA and protein levels were clearly correlated (cf. Fig. III.22 C, $R^2 = 0.40$ at log-log scale). Though this finding is consistent with earlier observations, the degree of correlation is higher than any previous large scale study in mammals (reviewed in: DE SOUSA ABREU *et al.,* 2009), indicating that parallel mRNA and protein quantification provides accurate data. Nevertheless, since the data is also affected by technical noise, the true correlation is expected to be higher. The coefficient of determination ($R^2 = 0.40$) indicates that at least 40% of variance in protein levels can be explained by differences in mRNA levels. Interestingly, this value is lower than in yeast (73%) (LU *et al.,* 2007), which may suggest a larger extent of post-transcriptional regulation occurring in mammalian cells.

In contrast to levels, a correlation between mRNA and protein half-lives was virtually absent (cf. Fig. III.22 D, $R^2 = 0.01$). Thus, many stable proteins have unstable mRNAs

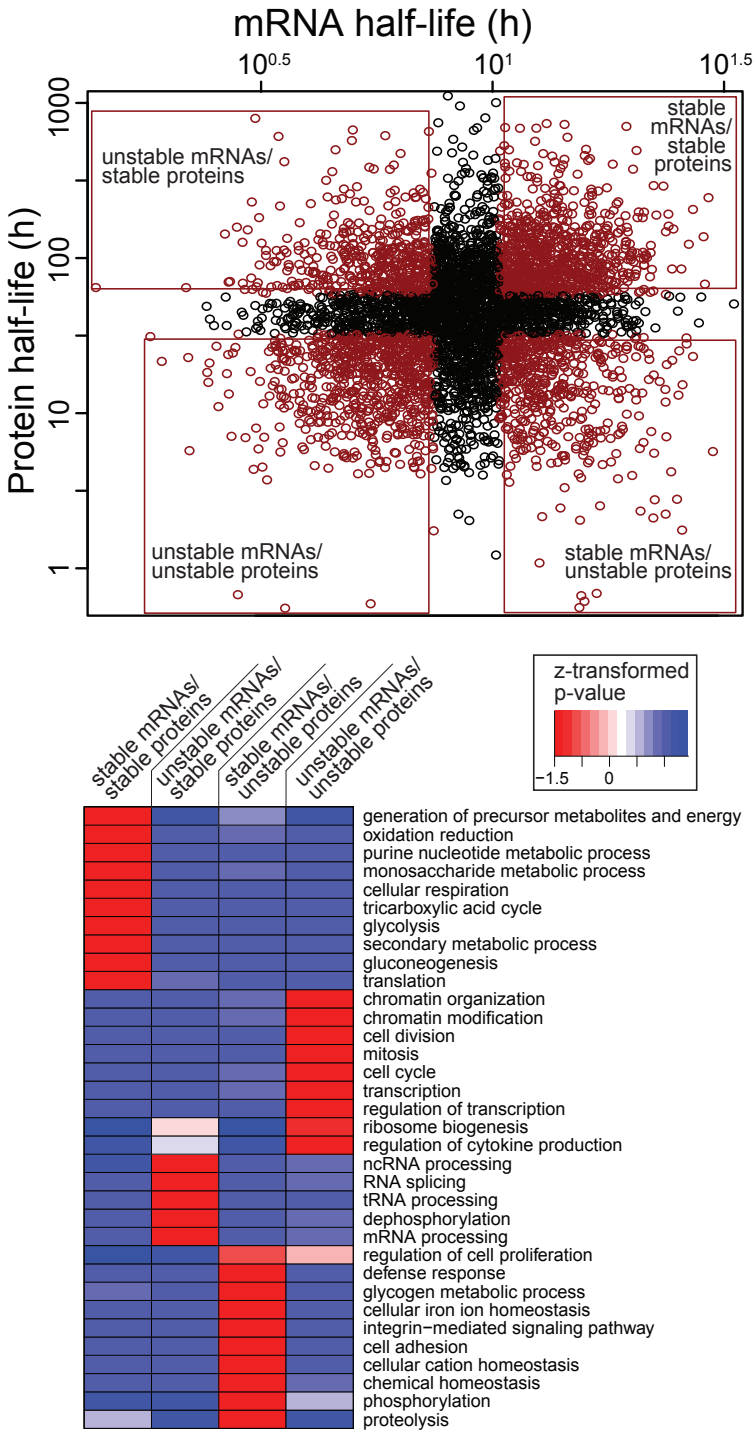


Fig. III.23 | Functional characteristics of genes with different mRNA and protein half-lives. Genes were grouped according to their combination of mRNA and protein half-lives and analyzed for enriched biological functions according to gene ontology. A heatmap was used to visualize hierarchical clustering of z-transformed p-values of significantly over-represented biological terms of genes with similar half-life characteristics.

and *vice versa.* To test whether this seemingly random distribution yet contains biologically relevant information, the plot was divided into certain regimes comprising gene subsets with different combinations of mRNA and protein half-lives. Altogether, four regimes were defined with distinct mRNA/protein half-lives distributions (cf. Fig. III.23, scatter plot and Material & Methods II.27) with (i) stable mRNAs and stable proteins, (ii) unstable mRNAs and unstable proteins, (iii) stable mRNAs but unstable proteins and finally (iv) unstable mRNAs but stable proteins. Next, the defined gene subsets were individually subjected to gene ontology analysis (Ashburner *et al.,* 2000) to examine whether a putative over-representation of functional terms (cf. Material & Methods II.27). P-values of enriched biological terms were z-transformed and organized and visualized using hierarchical clustering (cf. Fig. III.23, heatmap).

Genes with both stable mRNAs and proteins were primarily involved in constitutive cellular processes like translation (i.e. ribosomal proteins), respiration and central metabolism (glycolysis, citric acid cycle). Hence, many highly expressed "house-keeping" genes tended to encode stable mRNAs and proteins. In contrast, the subset with unstable mRNAs and proteins was strongly enriched for dynamically regulated genes such as transcription factors, chromatin modifying enzymes and genes with cell cycle-specific functions. Thus, many regulatory genes had low mRNA and protein half-lives. Intriguingly, genes with stable proteins but unstable mRNAs were involved in processing of mRNAs, tRNAs and non-coding RNAs. This suggests that many mammalian RNA-binding proteins are stable while their encoding transcripts tend to be short-lived, as noted recently for yeast (Mittal *et al.,* 2009). Finally, the set of genes with high mRNA and low protein stability was enriched in extracellular proteins. This is expected since secreted proteins have short cellular half-lives. Additionally, this subset contained proteins involved in cellular homeostasis, defense response and proteolysis.

In summary, the analysis showed that mRNA and protein half-lives are not independent. Instead, genes seem to have evolved specific combinations of mRNA and protein half-lives under functional constraints.

III.3.4.1 Half-lives substantially determine gene expression kinetics

mRNA and protein half-lives are important parameters that shape the dynamic properties of gene expression (Legewie *et al.*, 2008; Shalem *et al.,* 2008; Price *et al.,* 2010). Hence, some genes may have evolved specific combinations of mRNA and protein half-lives in order to respond expeditiously to an external signal. Along these lines, mRNA stability was recently shown to play a critical role in gene induction under several conditions in mammals (Hao & Baltimore *et al.,* 2009; Friedel *et al.,* 2009; Elkon *et al.,* 2010). However, the above presented findings regarding protein and mRNA half-lives (cf. III.3.4) suggest that also protein turnover plays a pivotal role in shaping the kinetics of gene expression. Protein half-lives were on average considerably longer than corresponding mRNA half-lives and, in addition, only marginally correlated to those. Thus, a measure with a predictive value for the kinetics of gene expression upon perturbations should ideally account for both, mRNA and protein stability.

When subjected to an abrupt change in the environmental condition, cells typically re-configure

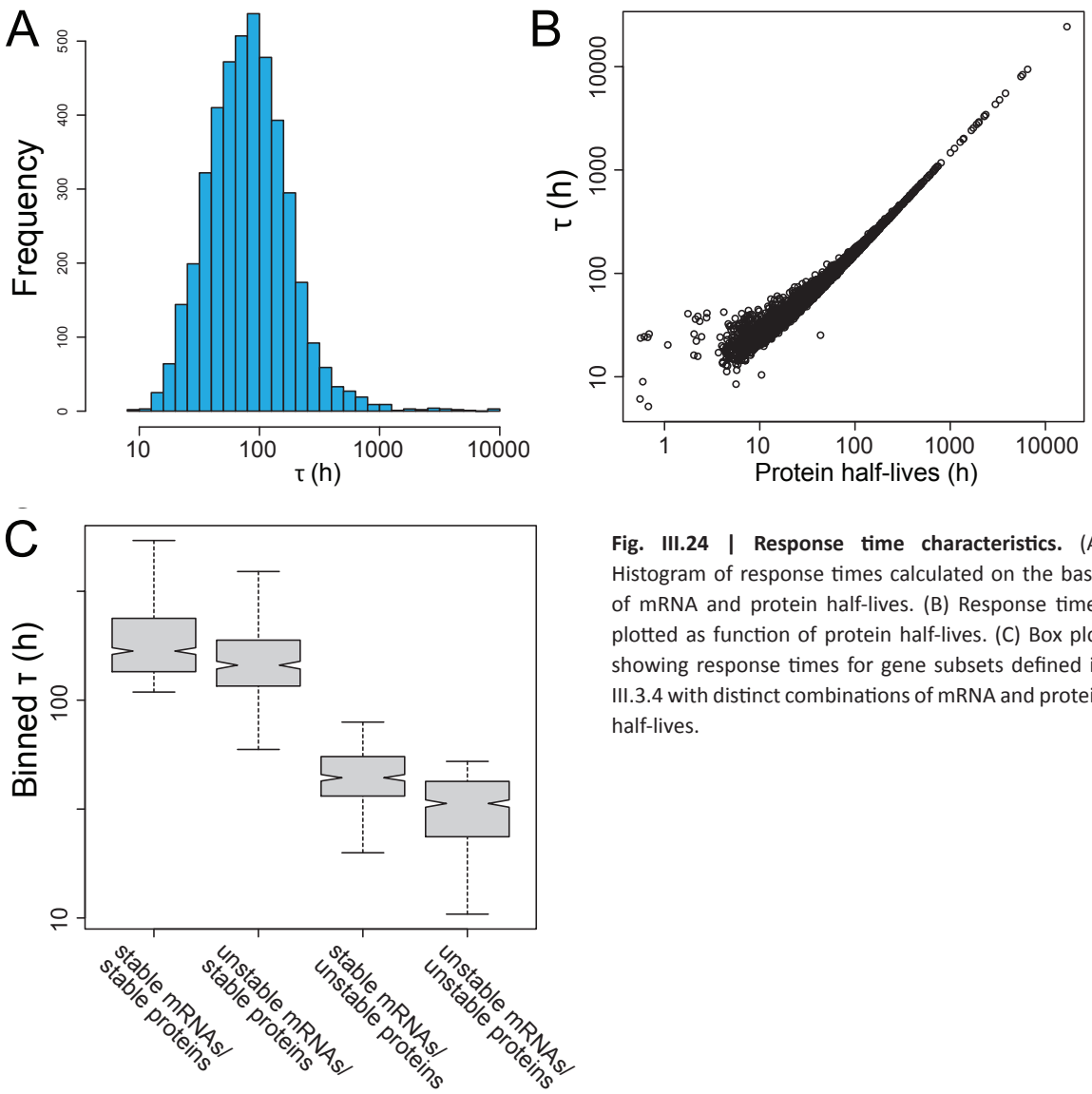


Fig. III.24 | Response time characteristics. (A) Histogram of response times calculated on the basis of mRNA and protein half-lives. (B) Response times plotted as function of protein half-lives. (C) Box plot showing response times for gene subsets defined in III.3.4 with distinct combinations of mRNA and protein half-lives.

gene expression followed by a transition into a new steady-state which is oftentimes different from the original "resting" state. The time that describes how fast protein levels relax towards the steady-state after perturbation can be mathematically expressed by so called response times τ (cf. Material & Methods II.29.3) (HEINRICH & SCHUSTER, 1996). Since response times are solely determined by mRNA and protein decay rates, they depend on both if the protein and the mRNA half-lives are of similar magnitude, while it is mainly set by the slowest decay rate in case mRNA and protein stability differ significantly from each other (LEGEWIE *et al.*, 2008).

The overall distribution of response times calculated on the basis of mRNA and protein half-lives in the present study is shown in Fig. III.24 A. Notably, the time-scale of response times was more similar to the time-scale of protein than mRNA half-lives (cf. Fig.III.22 A), indicating that protein turnover significantly influences the induction kinetics and therefore the temporal pattern of gene expression. Consistently, plotting response times as function of protein half-lives revealed that mRNA stability can only affect gene expression dynamics if protein stability is small, that is protein half-lives < ~ 20 h (cf. Fig. III.24 B).

In order to elucidate how the response times of genes relate to their biological functions,

median response times for the four combinations of mRNA and protein half-lives used in chapter III.3.7 were computed (cf. Fig. III.24 C). Both gene subsets with high protein stability (i.e. the subset with high mRNA and high protein stability and the subset with low mRNA but high protein stability) also displayed relatively high response times compared to regimes comprising genes with lower protein stability (cf. Fig. III.24 C). Hence, genes involved in "house-keeping" functions but also in various RNA-processing steps tend to have rather sluggish and inert gene expression kinetics (cf. Fig. III.23). As those genes can be expected to respond slowly and delayed upon environmental changes, flexibility in terms of a fast adaptation of gene expression of those proteins is most likely not critical to the survival of the organism. The converse was true for genes given by regimes containing genes involved in cell cycle control, transcription but also immune defense and phosphorylation (i.e. gene subsets with high mRNA and high protein turnover and gene subsets with low mRNA but high protein turnover, cf. Fig. III.23 and III.24 C). Those had relatively short response times which is largely attributable to their significantly shorter protein lifetimes. If in addition to short protein half-lives genes also had low mRNA stabilities, response times were shortest. In that case, rapidly induced changes in gene expression can obviously effectively transmit to affect the protein output.

Altogether, the results indicate that mRNA and protein half-lives have evolved under temporal constraints. Half-lives obviously optimize the expression kinetics of genes mediating the cellular response and adaptation to environmental changes with regard to temporal flexibility and rapidity.

III.3.4.2 Half-lives optimize gene expression towards energy constraints

In lower organisms, energy costs keep transcription and translation rates under selective pressure (Wagner, 2005). Energy consumption may therefore also be an important parameter that constrains mRNA and protein half-lives in mammals. Based on the acquired data, the theoretical energy required to synthesize mRNAs (i.e. primary transcripts) and proteins from their building blocks (nucleotide monophosphates and amino acids, respectively) were calculated in terms of high energy phosphates (cf. Material & Methods II.30). Notably, this calculation is naturally an oversimplification since the energy consumption of other cellular processes like nucleotide and amino acid synthesis, splicing and transport is not known in fibroblasts. Due to the high dynamic range of protein levels, overall energy costs (summed costs for mRNA and protein synthesis) were primarily determined by protein translation. Specifically, protein synthesis consumed about 95% of the energy budget while only ~ 5% was needed for transcription (data not shown).

Given the high costs of protein synthesis it was obvious to ask whether this process is optimized to minimize cellular energy consumption. Energy costs for protein translation depend on protein abundance, half-life and length with abundance being the most important factor since protein levels showed the highest dynamic range (cf. Fig. III.22 B). Thus, abundant proteins should have in general higher half-lives than less abundant ones to minimize cellular energy consumption. To this end, proteins were divided into three abundance classes and analyzed

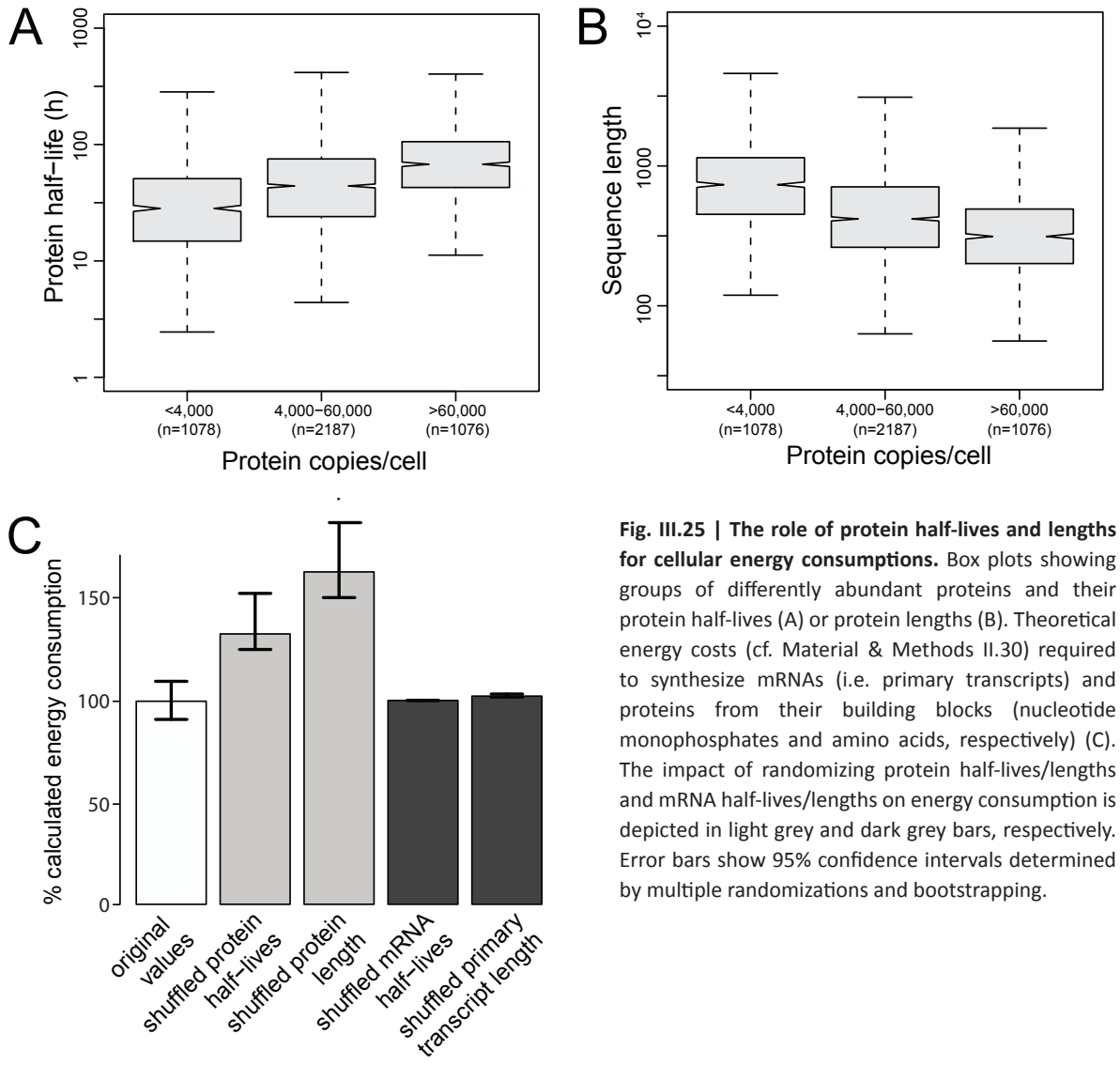


Fig. III.25 | The role of protein half-lives and lengths for cellular energy consumptions. Box plots showing groups of differently abundant proteins and their protein half-lives (A) or protein lengths (B). Theoretical energy costs (cf. Material & Methods II.30) required to synthesize mRNAs (i.e. primary transcripts) and proteins from their building blocks (nucleotide monophosphates and amino acids, respectively) (C). The impact of randomizing protein half-lives/lengths and mRNA half-lives/lengths on energy consumption is depicted in light grey and dark grey bars, respectively. Error bars show 95% confidence intervals determined by multiple randomizations and bootstrapping.

with regard to their protein half-lives and protein sequence lengths. The analysis revealed that abundant proteins were indeed significantly more stable than less abundant ones (cf. Fig. III.25 A, $p<10^{-15}$, Wilcoxon test). Likewise, abundant proteins were also significantly shorter (cf. Fig. III.25 B, $p<10^{-15}$). These observations are consistent with the hypothesis that mammalian protein half-lives and lengths evolved under energy constrains. In order to prove this hypothesis, all parameters were individually shuffled (cf. Fig. III.25 C). Remarkably, shuffling protein half-lives and lengths markedly increased theoretical energy consumption, whereas randomizing mRNA half-lives or lengths of primary transcripts had almost no impact. According to the functional analysis performed above (cf. III.3.4.1), predominantly abundant “housekeeping” genes tend to have stable mRNAs and proteins. The emerging global picture therefore is that most mRNAs and particularly proteins are designed to be rather stable unless they are involved in highly dynamically regulated cellular processes such as signal transduction. Due to the trade-off between dynamic regulation and energy efficiency this may represent an optimal design.

IV Discussion

IV.1. Establishment of pulsed SILAC (pSILAC)

Up to now, global methods for gene expression analysis have primarily focused on measuring changes at the mRNA level. However, the final step in expression of protein-coding genes is translation of mRNAs at the ribosome. In the present thesis, pulsed stable isotope labeling by amino acids in cell culture (pSILAC) with two heavy isotope labels has been established to directly quantify protein translation on a proteome-wide scale. The method can directly detect changes at the level of translation and, perhaps most importantly, irrespective of whether regulation is mediated by changes in mRNA stability (as seen for the transferrin receptor, TFRC, cf. Fig. III.4 C) or translation initiation (as shown for the ferritins, FTH1 and FTL, cf. Fig. III.4 B). While conventional microarray analysis can detect alterations in mRNA levels, the translationally regulated ferritins would have been most likely missed since corresponding transcript levels are known to be only marginally affected by iron treatment (Eisenstein & Munro, 1990). Given the fact that conservative estimates denote approximately one third of the expressed genes to be post-transcriptionally regulated (Ingolia *et al.,* 2009), the pSILAC approach is applicable to a wide range of biological questions (Mamane *et al.,* 2006; Valencia-Sanchez *et al.,* 2006). Combined with the microarray technique to measure changes at mRNA levels, the pSILAC approach should ultimately enable researchers to disentangle between regulation occurring at the level of mRNA stability and regulation occurring at the level of translation.

Recently, ribosome profiling has been suggested to measure actively transcribed mRNAs (Ingolia *et al.,* 2009). Here, following ribosome preparation and nuclease digestion, mRNA fragments protected from digestion by bound ribosomes are analyzed using next-generation sequencing technologies or conventional microarrays. Despite of its unparalleled precision in pinpointing translation start sites with nucleotide resolution, detecting and counting of ribosome-bound mRNAs does not necessarily imply that those transcripts are actively translated in the first place. For instance, binding of ribosomes to certain sequence elements like upstream open reading frames (uORF) was found to substantially affect translation as ribosomes may become stalled in the process of translation when encountering an uORF (Kozak, 1991; Morris & Geballe, 2000). In fact, uORFs were recently demonstrated to have a widespread, primarily negative impact on gene expression (Calvo *et al.,* 2008). This clearly challenges the capability of ribosome profiling to exclusively measure mRNAs in the process of translation. Unlike ribosome profiling, pSILAC detects changes in protein production at the very end of all steps in protein translation thus covering post-transcriptional regulation given by uORFs. On top of that even regulation occurring post-translationally like coordinated proteolysis mediated by post-translational modifications (PTMs) such as ubiquitinylation is captured (Hershko & Ciechanover, 1992).

While sequencing-based methods like ribosome profiling can basically cover the whole "translatome", though not without inconsistencies as mentioned above, any stable-isotope based quantification approach seeking a comprehensive, systems-wide analysis relies on sufficient mass spectrometry sampling depth. Hence, a major challenge in pSILAC is the detection of newly-synthesized proteins made against a bewildering background of pre-existing proteins. This issue applies especially to relatively short labeling times where small amounts of labeled, newly-synthesized proteins have to be traced in the presence of large amounts of unlabeled counterparts. In this case, the detection capability is severely limited by the dynamic range of the mass spectrometer. This limitation was also apparent from the kinetic of luciferase induction measured either by pSILAC or classic luminescence (cf. Fig. III.2 C) readout to assess and compare quantitative features of pSILAC. While overall correlation between the two approaches was excellent (cf. Fig. III.2 D; Pearson R = 0.99), marked errors bars of pSILAC measurements shortly after luciferase induction (approximately until 2 h post-induction) indicated variation in quantification of luciferase peptides. This observation points towards an intrinsic constraint of the pSILAC approach given by limited traceability of very small amounts of newly-synthesized proteins. Consequently, pSILAC requires labeling times of several hours to achieve decent amounts of newly-produced proteins which can hamper identification and quantification of transient changes in protein expression following perturbations. Yet pSILAC is not generally "blind" for newly-synthesized proteins from short labeling times. For instance, high turnover proteins can indeed produce decent amounts of newly-synthesized peptides within short time periods that are readily detectable by mass spectrometry. For instance, the heavy-labeled form of the Ribonucleoside-diphosphate reductase subunit M (Rrm2) for which the protein half-life has been determined in chapter III.3 was already detectable via mass spectrometry after a labeling time of 1.5 h (cf. Fig. III.14 A).

Monitoring luciferase induction by pSILAC led to another interesting observation in that a small amount of luciferase-derived peptides in the L (light) form (cf. Fig. III.2 A, B) could be detected. As luciferase was not expressed while cells were cultivated in the light medium, this peak most likely originates from recycling of light amino acids after transfer of the cells to M (medium-heavy) or H (heavy) medium. Alternatively, this peak could derive from remaining light amino acids in the cellular precursor pool (e.g. aminoacyl-tRNAs), a pervasive problem in stable isotope labeling of metazoans (Reeds & David, 1992; Hellerstein, 2004). Lastly, a more trivial explanation for the small L peak could be that the pTET promoter was to some degree "leaky". In this context, one cannot rule out that the promoter activity has been induced through traces of tetracycline or derivatives thereof in the culture medium. Since antibiotics are widely used in factory farming, it is conceivable that small amounts have been introduced to the culture medium by usage of fetal calf serum (FCS). However, in the pSILAC approach presented here this does not impair the accurate measurement of protein production for two reasons: First, the L peak is small (less than 10% of the heavy peaks after 10 h of induction), indicating that only a small fraction of newly synthesized peptides contain recycled and/or remaining light precursors. Second, and more importantly, light amino acids are incorporated

to the same extent in both samples. Hence, the abundance ratio of H versus M peptides will not be affected by incorporation of light amino acids.

Alternative strategies to study *de novo* protein synthesis employ selective labeling and enrichment of newly-synthesized proteins. In recent efforts, non-natural amino acids have been used to distinguish between newly-synthesized and pre-existing proteins. In the bioorthogonal non-canonical amino acid tagging approach termed BONCAT (Dieterich *et al.,* 2006), cells are grown on medium supplemented with azidohomoalanine (AHA) as a substitute for methionine to metabolically label newly-synthesized proteins. In the subsequent "click-chemistry" reaction, azide-labeled proteins are covalently coupled to an alkyne-bearing affinity tag enabling specific isolation of newly-produced proteins carrying such artificial amino acids.

Specific enrichment of newly-synthesized proteins as performed in BONCAT can indeed substantially facilitate their analysis. In doing so, sample complexity is decreasing what basically raises the chances of identifying them from a large pool of pre-existing proteins. However, in order to compare newly-synthesized proteins made under perturbed cellular conditions compared to untreated control cells, two separate populations of newly synthesized proteins are isolated in BONCAT and separately analyzed by mass spectrometry. Consequently, information gained by listing proteins identified in one against another cell population is limited to a qualitative description. In contrast, pSILAC does not only allow for direct qualitative but also for precise quantitative comparison of protein production between cellular conditions. While BONCAT cannot make a point about single proteins identified in one but not in the other population due to a stochastic sampling by the mass spectrometer, additional quantitative information obtained in pSILAC can considerably strengthen experimental validity. Lastly, the methionine substitute AHA used in BONCAT was found to induce growth arrest of *E.coli* during prolonged labeling (> 30 min) further limiting the versatility of the approach (Kramer *et al.,* 2009). Thus, BONCAT and pSILAC should rather be seen as complementary approaches. Given the increase in dynamic range and sensitivity as a result of a lower sample complexity, BONCAT may be preferentially used to detect transient changes in protein production. In contrast, pSILAC is better suited for investigation of biological processes altering protein synthesis on the hours time-scale. Here, owing to high accuracy in quantification, pSILAC can even determine subtle differences in protein production that are most likely not amenable to other methods.

Finally, it is worth mentioning that cells usually not compatible with conventional SILAC labeling can also be approached with pSILAC which is due to the fact that pSILAC does not require full proteome labeling. For instance, in a collaboration project dealing with translational regulation exerted by microRNAs, B-cells from context-specific knockout and control mice were isolated and cultured in pSILAC medium for a few days (data not shown). It turned out that such an *ex vivo* labeling of cells, though proceeding slower compared to standard cell lines due to a slower cell division, was sufficient to work out changes in protein production. Hence, the pSILAC approach is not limited to mere analysis of immortalized, standard cell lines.

In summary, the pSILAC method presented here to globally measure relative changes in protein synthesis across different cellular conditions is characterized by a very good dynamic range (cf.

Fig. III.2 C), high accuracy in quantification (cf. Fig. III.2 D) and very good reproducibility (cf. Fig. III.5). Hence, pSILAC offers a very good trade-off performance for the demand to measure translation on a global scale with a resolution close to methods working on a single-protein basis. Further improvements in MS instrumentation and software will most likely enable comprehensive measurement of the entire "translatome" in the near future (Cox & Mann, 2007).

IV.2. The impact of microRNAs on the proteome

In recent years, microRNAs (miRNAs) have emerged as a potent class of small regulatory RNAs exerting their repressive function by destabilization of target mRNAs and/or by inferring with the translation initiation or elongation step. Identifying functionally important miRNA targets is therefore key towards a comprehensive understanding of miRNA functions, for example for diseases based on oncogenic miRNAs termed "oncomirs" (Esquela-Kerscher & Slack, 2006). However, the above mentioned dual functionality of miRNAs makes experimental target identification particularly challenging. Based on the molecular mechanism of target recognition three main branches of large scale target identification have emerged: RISC (RNA-induced silencing complex, cf. Fig. I.6) pulldown experiments, transcriptomics and proteomics.

In RISC pulldown experiments the miRNA loaded RISC-complex is affinity purified together with associated mRNAs which are subsequently analyzed by microarray (Hock *et al.,* 2007; Karginov *et al.,* 2007). Such experiments provide information about the physical interaction between endogenous miRNAs and their targets. However, it is unclear if this physical interaction mediates target repression in all cases. Moreover, no information about the degree of target repression is provided.

Mainly because of difficulties in large scale protein quantification (Gygi *et al.,* 1999; Nie *et al.,* 2006; Tian *et al.,* 2004), mRNA expression profiling has so far been widely used as a surrogate for the identification of miRNA targets (Lim *et al.,* 2005). However, translationally regulated targets not susceptible to mRNA degradation would be missed.

Arguably, the most direct way of miRNA target identification is measuring changes at the protein level upon over-expression or depletion of certain microRNAs by quantitative proteomics. Proteins are the end-products of most genes and changes in protein levels naturally have a high functional relevance. Quantifying the impact of microRNAs at the protein level has the big advantage that it captures targets regulated by both mRNA degradation and translational repression. In a recent conceptually very similar study researchers employed mass-spectrometry based quantitative proteomics to assess the impact of miRNAs onto the proteome (Baek *et al.,* 2008). The group used the traditional SILAC method to measure how miRNAs affect steady-state levels of proteins two days post-transfection of specific microRNAs. In contrast, in the present study a novel variant of the SILAC method termed pulsed SILAC (pSILAC, cf. III.1) was used to globally measure changes in cellular protein synthesis in response to miRNA over-expression or knock-down (Selbach *et al.*, 2008). The pSILAC approach is independent of

individual protein turnover rates which can skew the effect of examined microRNAs onto the proteome (cf. Introduction I.2.2). In order to elucidate how much regulation is also reflected on the transcriptome, changes in mRNA levels employing conventional microarrays were monitored in parallel.

In addition to what is already known for mRNAs levels from a previous study (Lim *et al.,* 2005), the present findings clearly show that single miRNAs can affect protein production of hundreds of proteins (cf. Fig. III.9 A). Closer inspection of down-regulated proteins revealed that the presence of seed motifs located in the 3´UTRs of mRNAs strongly correlated with the observed fold changes, i.e. with down-regulation of protein production (cf. Fig. III.9 B, C, D and Fig. III.10 A). As the same held true for changes in mRNA levels (cf. Fig. III.11 C), the seed sequence appears as the primary motif directing both mRNA degradation as well as translational repression. In contrast, no seed enrichment was observed for up-regulated genes, indicating that the recently reported miRNA mediated activation of gene expression did not occur under the applied experimental conditions (Vasudevan *et al.,* 2007).

Intriguingly, the effects of miRNAs on proteins were overall quite modest as changes in their expression levels were on average less than twofold. As pSILAC exclusively measures changes in protein production, the finding cannot be explained by effects resulting from regulated turnover. Indeed, the observed pSILAC fold changes were on average even higher as fold changes measured via classic SILAC upon altering cellular miRNA expression levels (Baek *et al.*, 2008). This suggests that the analysis of microRNA-mediated regulation by classic SILAC was presumably compromised by protein turnover and that pSILAC is the more appropriate approach to study the impact of microRNAs onto the proteome.

In order to independently validate changes in protein production measured by pSILAC, luciferase reporter assays were performed with seed-carrying, endogenous 3´UTRs of nine mRNAs (Fig. III.10 B). Measuring of luciferase activity is one of the best quantitative methods currently available to study translational regulation since the method possesses high sensitivity and yields linear results over several orders of magnitude (Dyer *et al.,* 2000). Notably, a very good correlation of luciferase quantification with pSILAC ratios was observed despite the fact that the experimental procedures differ substantially and experiments were not done in parallel. In particular, there was no general trend of pSILAC to over- or underestimate changes in protein translation. However, other investigators observed much higher fold-changes (up to 30-fold) in a similar system (dsRNA transfection in HeLa cells) with artificial reporter constructs (Doench *et al.*, 2003). One explanation for this discrepancy is that very few (< 0.5%) 3'UTRs in the data set have more than three seed sites for a given miRNA, and this value is representative for the whole genome. Artificial reporter constructs are, however, designed to contain up to six closely spaced miRNA binding sites instead of entire, endogenous 3´UTRs with very few seed sites only. Hence, if a target has several seed sites, the repressive effect is multiplicative as has been observed in small-scale studies (Doench *et al.,* 2003; Rajewsky 2006). Remarkably, this cooperative effect is also apparent from the pSILAC fold-changes as the number of seeds per protein is inversely correlated with the magnitude of down-regulation (cf. Fig. III.11 C).

The parallel analysis of changes in mRNA levels by microarrays showed that repression of protein production is frequently mirrored by decreased transcript levels suggesting that most targets can be identified with the more cost-effective and straightforward conventional microarray technique (cf. Fig. III.11 B). However, by directly measuring changes in protein production, pSILAC data are *per se* more relevant to the phenotypes than microarray data. Notably, as transcript abundances were monitored before (t_1) and after (t_2) the pulse-labeling, it was found that for basically all examined miRNAs the correlation between mRNA fold-changes and pSILAC fold-changes became better at t_2 (cf. Fig. III.11 A and B and Supplementary Figure VII.2). This observation is subject of an ongoing, very controversial debate about microRNA function. It is still unclear whether mRNA degradation is an independent mechanism exerted by miRNAs or rather the consequence of the actual translational repression (Eulalio *et al.,* 2008). Though compelling evidence indicates that miRNA-mediated mRNA decay can in fact be uncoupled from translation inhibition (Wu *et al.,* 2006, Mishima *et al.,* 2006; Wakiyama *et al.,* 2007), the present findings at least suggest a temporal arrangement of the two effects where translational repression seems to precede mRNA degradation. However, since many reports convincingly demonstrate that the degree of relative contributions of translational repression and mRNA decay can considerably depend on structural features and the sequence context of individual miRNA:mRNAs hybrids (Grimson *et al.,* 2007), this observation should not be over-interpreted.

After all, the pSILAC approach identified a number of targets where translational repression obviously is the major component of mRNA silencing. Since corresponding changes at the mRNA level were barely detectable, such targets would most likely go unnoticed by traditional transcriptomics approaches. For instance, the well-known protein Dicer which is crucial for processing of virtually all miRNAs and silencing RNAs (siRNAs) was found to be primarily regulated translationally. To be precise, the protein that carries several let-7b 3´UTRs seeds was one of the most strongly up-regulated targets in the let-7b knock-down experiment with a > 4-fold change in the pSILAC but not in the microarray data (< 1.3 fold). Therefore, Dicer is a likely direct translational target of the microRNA let-7b.

In order to elucidate whether targets with a pronounced translational repression have functional properties in common, a subset of proteins with large differences in protein and mRNA fold-change (log2 fold-change pSILAC - mRNA < -0.3) were subjected to gene ontology analysis. Intriguingly, an over-representation of proteins synthesized at endoplasmic reticulum-(ER) associated ribosomes (GO categories: “intrinsic to membrane”, “endoplasmic reticulum”) was found. Evidently, translational repression appears to be stronger for mRNAs translated at ER-associated ribosomes as opposed to free cytosolic ribosomes. It is tempting to speculate that mRNAs from free ribosomes but not from ER-associated ribosomes are targeted to so called processing bodies (p-bodies) for degradation (Eulalio, 2007). As the ER is considered to lack proteolytic activity, this finding also suggests that translational repression may indeed be first in miRNA-mediated regulation of gene expression for that subset of targets (Nottrott *et al.,* 2006).

Apart from the above mentioned detection of Dicer as a translational target, the knock-down of let-7b as a highly abundant endogenous miRNA in HeLa cells was also in a different aspect highly informative. Since transfections of large amounts of miRNAs represent a considerable manipulation of cellular conditions, it was not clear to what extent the results are actually physiologically relevant. Adverse effects upon ectopic over-expression of microRNAs may arise from a competition in RISC loading between over-expressed and endogenous miRNAs that could potentially affect the results. In this regard, two scenarios are conceivable:

(i) Over-expressed miRNAs are not efficiently loaded into RISC, and the changes on protein production might therefore be underestimated.

(ii) Over-expressed miRNAs are loaded into RISC to very high levels which could lead to unspecific repression of proteins which are not targets under physiological miRNA concentrations.

The knock-down experiment, however, revealed that basically the same set of proteins down-regulated in the let7b over-expression experiment due to complementary seed sites was found to be de-repressed following let-7b knock-down (cf. Fig. III.12 A and B). Thus, the inverse modulation in protein production strongly suggests that artificial over-expression does not result in non-physiological effects. Except for the fact that pSILAC fold changes were overall marginally higher in the let-7b over-expression compared to the knock-down experiment, the presented data likely contain many physiologically relevant direct targets. Remarkably, the observed anti-correlation also held true for most of the ~2,700 proteins quantified in the experiments and not solely for the let-7b targets with seed-complementary sites, demonstrating that altering stationary levels of an endogenously expressed miRNA can affect the expression of thousands of genes (cf. Fig. III.12 A). It is conceivable that let-7b controls by targeting Dicer the maturation of diverse miRNAs which in turn affects a plethora of genes resulting in the observed proteome-wide changes. Yet a future challenge will be to work out genes that are directly targeted by miRNAs without bearing any canonical seed sites as already demonstrated for the transcription factor E2F2 and CDNK4 targeted by miR-24 and the miR-34 family, respectively (Lal *et al.,* 2009; Hermeking, 2010).

With one miRNA inducing expression changes in hundreds of proteins, it remains an open question how many of these changes are important to the function of the individual miRNA. Despite of carrying miRNA-specific seed-sites, it cannot be ruled out that some, or perhaps many of the proteins found to be mildly regulated are actually neutral due to tissue-specificity of microRNA targets (Lagos-Quintana *et al.*, 2002). Thus, it will be particularly illuminating to learn whether the proteins that are more strongly regulated have greater relevance to miRNA function, consistent with a switch-like model of miRNA activity (Reinhart *et al.,* 2000; van Rooij *et al.,* 2007; Liu & Olson, 2010). However, the observed wide-spread but mild changes induced by single miRNAs paint a picture of miRNAs as potent regulators that globally fine-tune protein expression by subtly affecting many targets simultaneously (Hobert, 2007).

IV.3. Dynamic properties of mammalian gene expression

Gene expression is a dynamic, multistep process that involves transcription, translation and turnover of mRNAs and proteins. All processes are subject to precisely controlled regulation to adapt gene expression to changing conditions occurring inside and outside the organism. Towards a quantitative description of how the combined effects of dynamic properties shape gene expression, large scale measurements of dynamic parameters are essential. Though methods to comprehensively record mRNA levels and mRNA degradation have become available in recent years, the knowledge about the relationship between mRNA and protein expression remains fragmentary. Due to extensive translational and post-translational regulation, differences in protein levels can be only partially explained by differences in mRNA levels resulting from transcription and decay. Recent efforts employing high-throughput technologies approached the connection between cellular transcripts and proteins by measuring steady-state concentrations. While most reports comparing mRNA and protein levels arrived at the conclusion that overall correlation is poor (reviewed in: DE SOUSA ABREU *et al.,* 2009; MAIER *et al.,* 2009), they cannot refer the discrepancy to a cause. Apart from the fact that most studies are still limited to a few hundred genes and that observations derived from prokaryotes cannot be simply transferred to eukaryotes, the actual limitation is given by the measurement of steady-state concentrations. mRNA and protein levels constitute the net outcome from coupled processes of synthesis and degradation and can hence not alone provide information about the underlying dynamic properties of gene expression. Alternatively, mRNA and protein turnover have been separately investigated with either drugs inhibiting transcription or translation or by the use of artificial fusion proteins (YANG *et al.*, 2003; BELLE *et al.,* 2006; YEN *et al.,* 2008). However, both methods are intrinsically suboptimal since the use of drugs and fusion proteins comes not without costs given by severe side effects or by affecting protein stability, respectively.

To overcome these potential shortcomings, in the present study both mRNA and protein half-lives as well as mRNA and protein concentrations were quantified simultaneously in mammalian cells on a genome-wide scale. For this to happen, stable isotope labeling by amino acids in cell culture (SILAC) has been combined with 4sU labeling, resulting in protein and mRNA turnover times and levels for more than 5,000 genes. Since all parameters were derived from the exact same data set and hence from the very same set of cells, error-prone steps given by integrating quantities from different sources were avoided.

Proteins were pulse-labeled with heavy amino acids for different lengths of time to cover a reasonable time range of protein turnover known to span several orders of magnitude in yeast (BELLE *et al.,* 2006). In addition, experiences about the time-scale protein production gained by stimulation of HeLa cells with iron turned out to be beneficial (cf. III.1.4). For instance, translational up-regulation of the ferritins took at least 4-6 h to arrive at H/L peptide peak ratios readily detectable by the mass spectrometer (cf. Fig. III.5). Thus, in order to capture both

high and low turnover proteins a time frame for labeling spanning more than 12 h was chosen. While the bulk of proteins were independently quantified at all three time points, only about 100 proteins were quantified at one time point only. Hence, only few proteins had extreme ratios, e.g. very high H/L ratios at the first time point or very small ratios only detectable at the last time point, suggesting that the labeling time was suitable to measure protein turnover in mammalian cells.

In order to account for all ratio measurements for half-life calculations, linear regression was applied to all H/L ratios obtained for individual proteins. As a quality test for the linear fit, the coefficient of termination R^2 was used. In total, the fit did satisfy the relatively conservative criterion of $R^2 \geq 0.9$ for 4,906 proteins, indicating that protein turnover could be accurately recorded by SILAC-based quantitative mass spectrometry. As expected for very unstable proteins, the H/L ratios were relatively high at the first time point (1.5 h) and then "saturated" at the later time points, that is, the H/L ratios were no longer increasing linearly. This finding points towards an intrinsic constraint of mass spectrometry to accurately capture extreme ratios with insufficient signal-to-noise ratios. For instance, when applying the linear regression to all three H/L ratios obtained for Cyclin D1 (1.4 at t_1, 4.9 at t_2 and 7.6 at t_3) the protein half-life was first underestimated with 4.5 h. Hence, for proteins with an H/L ratio ≥ 0.5 at the first time point (1.5 h) only this ratio was taken for calculation of protein turnover. In doing so, the half-life of Cyclin D1 was determined to be about 1.2 h, which could also be confirmed by cylcoheximide-chase experiments.

Robust quantification, however, relies not only on the instruments capability to precisely measure H/L ratios, it also necessitates that the ratios reliably reflect protein synthesis. Protein turnover can in particular be under-estimated due to the presence of light (L) amino acids that remain in the cell´s precursor pool after transferring them to heavy (H) SILAC medium, e.g. by recycling of light proteins. However, closer inspection of missed-cleaved, heavy peptides revealed that the measured protein half-lives were not appreciably affected by non-uniform incorporation of heavy amino acids. This is advantageous over *in vivo* experiments that often require complex corrections for different amino acid precursor pools (Doherty *et al.,* 2005). Unlike cells in culture, where the starting biomass is rather small, intact animals usually provide a substantial, tissue-specific pre-existing source of amino acids that enters the precursor pool. Nonetheless, the molecular basis for the outstanding regenerative capabilities of the urodele amphibian *Notophthalmus viridescens* has recently been investigated in a tissue-specific context (the authors restricted their analysis to tail tissue) with a pulsed *in vivo* SILAC approach (Looso *et al.,* 2010).

Protein half-lives measured in NIH3T3 cells with the SILAC approach were independently validated by cycloheximide-chase experiments. It should be mentioned that this kind of analysis only allows for a rough classification of protein stability. In this type of experiment cellular protein degradation is monitored following inhibition of protein synthesis by cycloheximide. However, those toxic conditions potentially distort the measurements. Monitoring the decay of proteins over many hours can therefore be error-prone, since cells will likely become

apoptotic. Despite of those potential limitations, the results obtained by the cycloheximide approach were in overall very good accordance with SILAC measurements (cf. Fig. III.16), thereby further confirming the validity of the approach. Ultimately, traditional radioactive pulse-chase experiments that enable a precise, though cumbersome determination of half-lives can be used to validate individual SILAC half-lives (Kenney, 1967).

Transcript half-lives were obtained essentially as described previously (Doelken *et al.,* 2008) using 4sU to metabolically label and isolate newly-synthesized RNA from total RNA. Comparison of half-lives measured for NIH3T3 in this study compared to those already publicly available revealed a strong correlation on a log-log scale (Pearson R = 0.64). This cannot necessarily be expected given the fact that different cell clones, and more importantly, different technologies (microarray vs. next-generation sequencing) were used to measure the distinct RNA fractions (i.e. newly-synthesized, pre-existing and total RNA). The determination of mRNA half-lives with the 4sU approach requires that newly-synthesized transcripts are separated from total RNA with equal efficiencies. However, transcripts containing large numbers of uridines that can be replaced by 4sU are inherently more prone to isolation by affinity beads. This bias became apparent when plotting the ratio of newly-synthesized/total RNA as a function of the number of uridines per transcript (cf. Fig. II.1 A). Obviously, the turnover of transcripts containing less uridines were significantly under-estimated. In order to alleviate the bias, LOWESS regression was applied to the ratios prior to half-life calculations (cf. Fig. II.1 B).

A comparison of mRNA and protein half-lives revealed that proteins were on average five times more stable and also spanned a higher dynamic range than mRNA half-lives (cf. Fig. III.22 A). Plotting mRNA and protein lifetimes against each other demonstrated that a correlation was virtually absent. Yet it seemed likely that this seemingly chaotic distribution follows certain design principles. Intriguingly, the analysis of genes with certain combinations of mRNA and protein half-lives revealed that they share common biological functions (cf. III.3.4), suggesting that they evolved under similar constraints. Particularly informative in this regard was the fact that mainly genes involved in dynamically regulated processes such as transcription factors and cell cycle regulation were characterized by low mRNA and protein stability. This indicated that one of the underlying design principles shaping cellular half-lives may be the ability of cells to modulate gene expression of specific genes in a temporally highly flexible manner. Consistently, the analysis of response times as a measure for how fast protein levels relax towards the steady-state after perturbation revealed that expression levels of such genes can be rapidly adapted, suggesting that an exact expression timing is crucial for the survival of the entire organism. On the contrary, genes involved in constitutive processes (e.g. translation, respiration and central metabolism) were characterized by rather stable mRNAs and proteins and hence long response times.

By definition, response times can only be influenced by both mRNA and protein half-lives when they are of similar magnitude (cf. Material & Methods II.29.3), which was the case for the above mentioned dynamically regulated genes. Changes in gene expression occurring at the mRNA level, that is, by changes in transcription or mRNA degradation, can thus effectively

transmit to affect the protein output. It is therefore tempting to speculate that those genes are at least to some degree transcriptionally regulated. This is consistent with a recent small-scale study were mRNAs and proteins of transcriptionally induced signaling species were all unstable (Legewie *et al.,* 2008). Moreover, in two recent reports (Hao & Baltimore *et al.,* 2009; Elkon *et al.,* 2010), mRNA stability was found to have a major impact on the temporal order of genes encoding proteins involved in inflammatory processes. Amongst those genes that showed rapid changes in gene expression upon stimulation with TNFalpha, known transcriptionally regulated genes such as Fos (Proto-oncogene c-Fos), TTP (Tristetraproline) and JunB (Transcription factor jun-B) were found. Remarkably, a comparison of such genes with the data obtained in the present study revealed, that both mRNA and protein stability were of comparable magnitude and thus basically confer a transcriptional regulation. In contrast, the fact that also genes with high mRNA but low protein stability showed fast response kinetics points towards a different type of regulation (cf. III.3.4.1.). Although changes in transcription will hardly affect mRNA levels in a timely manner, a translational regulation that is independent of mRNA stability is conceivable. Notably, cellular homeostasis genes in this group may thus be regulated at the level of translation. In fact, this set contains two ferritins that are well-known to be up-regulated translationally in response to iron (Hentze & Muckenthaler, 2004).

Another putative design principle that became apparent from the functional analysis may be energy efficiency (Wagner, 2005). To this end, theoretical energy costs required to synthesize mRNAs (i.e. primary transcripts) and proteins from their building blocks (nucleotide monophosphates and amino acids, respectively) were calculated (cf. Material & Methods II.30). This calculation underestimates the real costs since the energy required for nucleotide synthesis, amino acid synthesis, splicing and transport are unknown in fibroblasts and therefore omitted. Yet the analysis has led to instructive insights. For instance, the relative costs needed for mRNA and protein synthesis were significantly lower than expected at random, that is, shuffling protein half-lives and protein lengths markedly increased theoretical energy consumption. Intriguingly, due to the high dynamic range of protein levels the contribution of mRNAs to overall energy costs (i.e. summed costs for mRNA and protein synthesis) was almost negligible. More precisely, protein synthesis consumed about 95% of the overall energy budget while only ~ 5% was needed for transcription (data not shown). Along these lines, it also turned out that 80% of the energy for translation is required to synthesize 20% of all proteins. Hence, protein synthesis follows the Pareto principle (“80/20 rule”) with a small fraction of proteins consuming most of the energy. If gene expression was optimized under energetic constraints abundant and therefore highly expensive proteins are expected to be more stable and shorter than less abundant ones. Indeed, abundant proteins were significantly more stable than less abundant ones (cf. Fig. III.25 A) and they were also significantly shorter (cf. Fig. III.25 B). The emerging global picture therefore is that most mRNAs and particularly proteins tend to be stable unless genes need to respond quickly to a stimulus. Due to the trade-off between dynamic regulation and energy efficiency this may represent an optimal design.

Despite of the undeniable role of mRNA and protein turnover in regulation of cellular function,

the basic molecular mechanisms driving degradation are far from being well understood. In contrast to mRNAs, where at least some molecular determinants (e.g. *cis*-regulatory sequence features as AU-rich elements and Pum2 binding sites, cf. III.3.2) have been identified to clearly affect transcript stability, clear-cut protein degradation signals (degrons) remain elusive so far. Although several putative degrons have been suggested (e.g. physical/sequential features such as number of residues, overall hydrophobicity, folding properties, intrinsically unstructured regions, PEST motifs, etc.), they were all found to serve as rather weak signals for intracellular protein degradation (reviewed in: Tompa *et al.,* 2008). It seems likely that protein degradation is not only determined by a single characteristic, but is a multi-factorial process that shows large protein-to-protein variations. Yet the (simplified) analysis of potentially unstructured amino acid regions performed in the present thesis suggested that the degree of disorder is inversely correlated with protein half-life (cf. Fig. III.20). Interestingly, several studies performing genome-wide predictions of protein disorder came to the conclusion that intrinsically unstructured proteins predominantly carry out important functions in signal transduction and transcription regulation (Iakoucheva *et al.,* 2002; Ward *et al.,* 2004; Minezaki *et al.,* 2006). Given the high susceptibility of unstructured regions for the intracellular degradation machinery, this structural feature may represent a conceivable means and mechanism to confer dynamic gene expression (Dyson & Wright, 2005) of transcription factors, chromatin modifying enzymes and proteins with cell cycle-specific functions.

Another conceivable design principle became apparent from the observation that many mammalian RNA-binding proteins were stable but encoded by unstable transcripts, as also recently seen in yeast (Mittal *et al.,* 2009). Since many RNA-binding proteins bind their own messenger RNA, this observation is indicative of a post-transcriptional negative feedback-loop for RNA-binding proteins. Consistent with this idea an over-representation of RNA-binding protein motifs could be observed in high-turnover mRNAs (data not shown).

In addition to half-lives, parallel mRNA sequencing and mass spectrometry also allowed to estimate cellular mRNA and protein copy numbers. Transcript levels were calculated based on the number of sequencing reads in conjunction with information on cellular mRNA content as described before (Mortazavi *et al.,* 2008). For example, for the popular “housekeeping” mRNA encoding for of GAPDH 2,450 copies/cell were calculated. A previous study carefully determined GAPDH copy number in NIH3T3 cells by single-cell quantitative real-time PCR or bulk measurements (Marcus *et al.,* 2006). Intriguingly, the calculated value fell almost exactly between the two previously published values for bulk and single cell data (1,840 and 2,940 copies/cell, respectively). Despite this intriguing consistency, further experimental evidence confirming the calculated mRNA levels in the present study would certainly be beneficial. Absolute quantification using latest next-generation sequencing technologies is still in its infancy. Specifically, it remains not entirely clear whether the obtained RPKM values used for deduction of absolute concentrations truly scale linearly with absolute transcript amounts over several orders of magnitude. To address this, spiking known concentrations of reference transcripts in the measured sample covering a huge dynamic range (four to five orders of

magnitude) has been suggested (Mortazavi *et al.,* 2008). Alternatively, qPCR-based absolute quantification can be performed. The method is, however, impractical for validation of many transcripts and can be imprecise due to potentially error-prone steps (e.g. extraction of RNA of defined amount of cells, differences in primer binding efficiencies in RT-PCR, cloning issues, etc.) (Bustin, 2000).

Similarly, obtaining information about absolute protein copy numbers is also challenging. Absolute numbers can be obtained by mass spectrometry employing different approaches (Silva *et al.,* 2006; Lu *et al.,* 2007; Malmstrom *et al.,* 2009) which can be broadly speaking divided into two strategies. Quantification can be based on (i) spectral counting of peptides that neglect intensities as a possible source of information or (ii) can be achieved by extraction of intensities of a subset of peptides (e.g. the top three most intense ones) thereby neglecting all other peptide intensities obtained for a protein. To overcome these limitations, the sum of peak intensities of all peptides matching to a specific protein was used in a newly-devised approach termed iBAQ ("intensity-based absolute quantification"). When divided by the number of theoretically observable peptides, the resulting iBAQ values were shown to provide an accurate proxy for protein levels. Depending on the individual experiment to identify and quantify UPS2 proteins of known concentrations, iBAQ explained between 90 and 95% of the differences in absolute protein amounts (cf. Fig. III.21) and significantly outperformed all other approaches. To further test the method´s ability to quantify proteins expressed at vastly different cellular levels, yeast was subjected to absolute quantification because quantitative information based on tag-western-blots/GFP-fusion proteins and ribosome profiling is available for this model organism (Ghaemmaghami *et al.,* 2003; Ingolia *et al.,* 2009). Cellular copy numbers of 1,384 proteins determined by iBAQ ranged from less than 100 to more than 10^6 copies per cell (data not shown). The most abundant proteins were glycolytic enzymes, translation elongation factor and ribosomal proteins, consistent with a proteome optimized for fermentation and rapid growth. Overall, the data correlated surprisingly well with translation rates measured by ribosome profiling (cf. Supplementary Fig. VII.5, R^2=0.66). This correlation is considerably higher than the correlation between ribosome profiling and the tag-western-blot/GFP-fusion data from the same group (cf. Supplementary Fig. VII.5, R^2=0.29). The results indicate therefore that the relatively simple and straightforward iBAQ-based quantification was more reliable than values obtained by the laborious tag-western-blot/GFP-fusion experiments, which are still frequently used to globally compare mRNA and protein levels.

Despite a huge spread and a much higher dynamic range of proteins, cellular levels of mRNAs and proteins were clearly correlated (cf. Fig. III.22 C, R^2 = 0.40 at log-log-scale). Although this correlation is higher than in any previous mammalian study it is still much lower than in yeast (73%) (Lu *et al.,* 2007). Gene expression in higher eukaryotes may show a higher degree of post-transcriptional and post-translational regulation. Alternatively, it is also possible that the yeast system is special due to its extreme adaptation for rapid growth (Griffin *et al.,* 2002).

However, it should also be stressed that the data set presented in this study represents average values derived from a population of dividing, non-synchronized cells. Therefore, one cannot

make a point about half-lives and copy numbers in individual cells or at different cell cycle stages and cellular compartments. For instance, cyclins were found to have rather unstable transcripts and proteins. Since cyclins are regulated at the level of protein degradation by ubiquitin-mediated proteolysis in a cell cycle dependant manner, protein stabilities most likely undergo changes throughout the cell cycle (Bloom & Cross, 2007). Similarly, transcripts recruited to specific cellular compartments such as p-bodies or stress granules presumably have different stabilities compared to their cytoplasmic counterparts (reviewed in: Parker & Sheth, 2007. Also, at the single cell level, the role of protein degradation for protein abundance may be higher. In *E. coli*, single cell based methods revealed that protein levels fluctuate considerably (Rosenfeld *et al.,* 2005; Taniguchi *et al.,* 2010).

Altogether, the here presented data set is currently the largest and most comprehensive source of mRNA and protein half-lives and levels and thus provides *per se* a rich resource for the scientific community that can be mined in many ways that are beyond the scope of this study. For instance, protein copy numbers provide valuable information for modeling of cellular processes and can provide insights into the stoichiometry of protein complexes (Malmstrom *et al.,* 2009). Half-lives can be used to assess the impact of sequence features being under suspicion to destabilize mRNAs or proteins. The data may even prove useful for *de novo* discovery of yet unknown sequence elements or refinement of existing rules governing mRNA or protein degradation.

V Conclusions and Outlook

Most studies on global gene expression performed to date concentrated exclusively on measuring changes in steady-state mRNA levels. Although it is now clear that post-transcriptional and post-translational regulation play a central role, investigation of these events is still in its infancy (Keene, 2007; Komili & Silver, 2008). The pulsed SILAC approach presented in this thesis to explicitly study differences in protein translation is a valuable step in that direction. With the use of pSILAC the impact of microRNAs onto the proteome given by control of mRNA stability and protein translational has been qualitatively and quantitatively assessed. Other, not less important and exciting biological systems known to have a marked "translational dimension" are awaiting their examination. Amongst them, the study of mTOR signaling and hypoxic response but also the impact of RNA-binding proteins on gene expression deserve research efforts, in particular due to their crucial role in disease development.

The study of protein synthesis in the present thesis has also been harnessed to deduce fundamental dynamic parameters of mammalian gene expression. By combining SILAC with 4sU labeling in a pulsed fashion protein and mRNA half-lives and concentrations for more than 5,000 genes were obtained. This has led to the insight, that mRNA and protein half-lives share common biological functions and hence likely evolved under similar constraints. The four simultaneously measured parameters can now be directly used to quantitatively model mammalian gene expression dynamics, i.e. a comprehensive description of transcription, mRNA degradation, translation and protein degradation rates for thousands of genes is within reach.

In this context, it will be exciting to evaluate on a global level which factor, translation efficiency or protein half-life, dominates the correlation between mRNA and protein abundances. Without regulation at the post-transcriptional and post-translational level protein and mRNA abundance would correlate perfectly. In this study a coefficient of determination (R^2) between mRNA and protein copy numbers of 0.40 has been observed. If an absence of technical and biological noise is assumed, this would mean that ~40% of the variance in protein levels can be explained by differences in mRNA levels (given by transcription and mRNA decay). The remaining 60% are caused by different translation rates and protein stabilities. In yeast, one third of all transcripts is assumed to be translationally regulated with translation efficiencies spanning roughly two orders of magnitude (Ingolia *et al.,* 2008). In contrast, individual mouse protein half-lives measured in the present study ranged from a few minutes to more than hundred hours, a more than 1000-fold range which is also consistent with half-life data obtained in yeast (Belle *et al.*, 2006). Given the higher dynamic range of protein half-lives, it is tempting to speculate that protein turnover may affect the correlation between mRNA and protein abundances to a larger degree. Here mathematical models will help to ultimately explain the discrepancies observed between mRNAs and proteins. Additional methods like sequencing of nascent transcripts and ribosome profiling will add valuable information (Core *et al.,* 2008; Ingolia *et al.,* 2009). With the rapid evolution of genome-scale quantitative technologies, unraveling the principles of

transcriptional, post-transcriptional and post-translational regulation may eventually become feasible.

Despite of its appealing attempt to find answers to fundamental biological questions, all work in that direction also has important practical implications. For instance, RNA interference (RNAi) as a promising powerful therapeutic agent to silence pathogenic gene products requires a proper understanding of the kinetics of the process and the parameters that can affect the resulting gene silencing (Bartlett & Davis, 2006). Amongst such parameters, knowledge about target mRNA and protein stability is key to specifically tailor and optimize the treatment since it helps to assess to what degree and more importantly, when RNAi-induced knockdown phenotypes can be expected to become apparent. Notably, siRNA-mediated gene silencing can also result in unintended repression of genes other than the desired target, referred to as “off-target” effects (Jackson & Burchard, 2006). As this occurs in a manner reminiscent of target silencing by miRNAs, the pSILAC approach can in turn be utilized to study the extent of such side-effects on the proteome-level. This will facilitate the selection of siRNAs of high therapeutical relevance.

VI References

Aebersold, R. and M. Mann (2003). “Mass spectrometry-based proteomics.” Nature 422(6928): 198-207.

Andersen, J. S., Y. W. Lam, et al. (2005). “Nucleolar proteome dynamics.” Nature 433(7021): 77-83.

Ashburner, M., C. A. Ball, et al. (2000). “Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.” Nat Genet 25(1): 25-9.

Baek, D., J. Villen, et al. (2008). “The impact of microRNAs on protein output.” Nature 455(7209): 64-71.

Bantscheff, M., M. Schirle, et al. (2007). “Quantitative mass spectrometry in proteomics: a critical review.” Anal Bioanal Chem 389(4): 1017-31.

Barreau, C., L. Paillard, et al. (2005). “AU-rich elements and associated factors: are there unifying principles?” Nucleic Acids Res 33(22): 7138-50.

Belle, A., A. Tanay, et al. (2006). “Quantification of protein half-lives in the budding yeast proteome.” Proc Natl Acad Sci U S A 103(35): 13004-9.

Besse, F. and A. Ephrussi (2008). “Translational control of localized mRNAs: restricting protein synthesis in space and time.” Nat Rev Mol Cell Biol 9(12): 971-80.

Blagoev, B., S. E. Ong, et al. (2004). “Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics.” Nat Biotechnol 22(9): 1139-45.

Bloom, J. and F. R. Cross (2007). “Multiple levels of cyclin specificity in cell-cycle control.” Nat Rev Mol Cell Biol 8(2): 149-60.

Brodersen, P. and O. Voinnet (2009). “Revisiting the principles of microRNA target recognition and mode of action.” Nat Rev Mol Cell Biol 10(2): 141-148.

Bruggeman, F. J. and H. V. Westerhoff (2007). “The nature of systems biology.” Trends Microbiol 15(1): 45-50.

Bussemaker, H. J., H. Li, et al. (2001). “Regulatory element detection using correlation with expression.” Nat Genet 27(2): 167-71.

Bustin, S. A. (2000). “Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays.” J Mol Endocrinol 25(2): 169-93.

Calkhoven, C. F., C. Muller, et al. (2002). “Translational control of gene expression and disease.” Trends Mol Med 8(12): 577-83.

Calvo, S. E., D. J. Pagliarini, et al. (2009). “Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans.” Proc Natl Acad Sci U S A 106(18): 7507-12.

Canas, B., D. Lopez-Ferrer, et al. (2006). “Mass spectrometry technologies for proteomics.” Brief Funct Genomic Proteomic 4(4): 295-320.

Cascante, M., E. Melendez-Hevia, et al. (1995). “Control analysis of transit time for free and enzyme-bound metabolites: physiological and evolutionary significance of metabolic response times.” Biochem J 308 (Pt 3): 895-9.

Chang, T. C. and J. T. Mendell (2007). “microRNAs in vertebrate physiology and human disease.” Annu Rev Genomics Hum Genet 8: 215-39.

Cohen, A. A., N. Geva-Zatorsky, et al. (2008). “Dynamic proteomics of individual cancer cells in response to a drug.” Science 322(5907): 1511-6.

Core, L. J., J. J. Waterfall, et al. (2008). “Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.” Science 322(5909): 1845-8.

Cox, J. and M. Mann (2007). “Is proteomics the new genomics?” Cell 130(3): 395-8.

Cravatt, B. F., G. M. Simon, et al. (2007). “The biological impact of mass-spectrometry-based proteomics.” Nature 450(7172): 991-1000.

Desiderio, D. M. and M. Kai (1983). “Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue.” Biomed Mass Spectrom 10(8): 471-9.

Dice, J. F. (1987). “Molecular determinants of protein half-lives in eukaryotic cells.” FASEB J 1(5): 349-57.

Dieterich, D. C., A. J. Link, et al. (2006). “Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT).” Proc Natl Acad Sci U S A 103(25): 9482-7.

Doench, J. G., C. P. Petersen, et al. (2003). “siRNAs can function as miRNAs.” Genes Dev 17(4): 438-42.

Doherty, M. K., C. Whitehead, et al. (2005). “Proteome dynamics in complex organisms: using stable isotopes to monitor individual protein turnover rates.” Proteomics 5(2): 522-33.

Dolken, L., Z. Ruzsics, et al. (2008). “High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay.” RNA 14(9): 1959-72.

Domon, B. and R. Aebersold (2006). “Mass spectrometry and protein analysis.” Science 312(5771): 212-7.

Dyer, B. W., F. A. Ferrer, et al. (2000). “A noncommercial dual luciferase enzyme assay system for reporter gene analysis.” Anal Biochem 282(1): 158-61.

Dyson, H. J. and P. E. Wright (2005). “Intrinsically unstructured proteins and their functions.” Nat Rev Mol Cell Biol 6(3): 197-208.

Eisenstein, R. S. and H. N. Munro (1990). “Translational regulation of ferritin synthesis by iron.” Enzyme 44(1-4): 42-58.

Elias, J. E. and S. P. Gygi (2007). “Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.” Nat Methods 4(3): 207-14.

Esquela-Kerscher, A. and F. J. Slack (2006). “Oncomirs - microRNAs with a role in cancer.” Nat Rev Cancer 6(4): 259-69.

Eulalio, A., I. Behm-Ansmant, et al. (2007). “P bodies: at the crossroads of post-transcriptional pathways.” Nat Rev Mol Cell Biol 8(1): 9-22.

Eulalio, A., E. Huntzinger, et al. (2008). “Getting to the root of miRNA-mediated gene silencing.” Cell 132(1): 9-14.

Flynt, A. S. and E. C. Lai (2008). “Biological principles of microRNA-mediated regulation: shared themes amid diversity.” Nat Rev Genet 9(11): 831-42.

Gebauer, F. and M. W. Hentze (2004). “Molecular mechanisms of translational control.” Nat Rev Mol Cell Biol 5(10): 827-35.

Gerber, S. A., J. Rush, et al. (2003). “Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS.” Proc Natl Acad Sci U S A 100(12): 6940-5.

Ghaemmaghami, S., W. K. Huh, et al. (2003). “Global analysis of protein expression in yeast.” Nature 425(6959): 737-41.

Gossen, M. and H. Bujard (1992). “Tight control of gene expression in mammalian cells by tetracycline-responsive promoters.” Proc Natl Acad Sci U S A 89(12): 5547-51.

Greenbaum, D., C. Colangelo, et al. (2003). “Comparing protein abundance and mRNA expression levels on a genomic scale.” Genome Biol 4(9): 117.

Griffin, T. J., S. P. Gygi, et al. (2002). “Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae.” Mol Cell Proteomics 1(4): 323-33.

Grimson, A., K. K. Farh, et al. (2007). “MicroRNA targeting specificity in mammals: determinants beyond seed pairing.” Mol Cell 27(1): 91-105.

Gsponer, J., M. E. Futschik, et al. (2008). “Tight regulation of unstructured proteins: from transcript synthesis to protein degradation.” Science 322(5906): 1365-8.

Guerrera, I. C. and O. Kleiner (2005). “Application of mass spectrometry in proteomics.” Biosci Rep 25(1-2): 71-93.

Gygi, S. P., B. Rist, et al. (1999). “Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.” Nat Biotechnol 17(10): 994-9.

Gygi, S. P., Y. Rochon, et al. (1999). “Correlation between protein and mRNA abundance in yeast.” Mol Cell Biol 19(3): 1720-30.

Hafner, M., M. Landthaler, et al. (2010). “Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP.” Cell 141(1): 129-41.

Hampf, M. and M. Gossen (2007). “Promoter crosstalk effects on gene expression.” J Mol Biol 365(4): 911-20.

Hanke, S., H. Besir, et al. (2008). “Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level.” J Proteome Res 7(3): 1118-30.

He, L. and G. J. Hannon (2004). “MicroRNAs: small RNAs with a big role in gene regulation.” Nat Rev Genet 5(7): 522-31.

Heinrich, R., B. G. Neel, et al. (2002). “Mathematical models of protein kinase signal

transduction." Mol Cell 9(5): 957-70.

Heinrich, R. and T. A. Rapoport (1975). "Mathematical analysis of multienzyme systems. II. Steady state and transient control." Biosystems 7(1): 130-6.

Heinrich, R. and S. Schuster (1996). The regulation of cellular systems.

Hellerstein, M. K. (2004). "New stable isotope-mass spectrometric techniques for measuring fluxes through intact metabolic pathways in mammalian systems: introduction of moving pictures into functional genomics and biochemical phenotyping." Metab Eng 6(1): 85-100.

Hentze, M. W., M. U. Muckenthaler, et al. (2004). "Balancing acts: molecular control of mammalian iron metabolism." Cell 117(3): 285-97.

Hermeking, H. "The miR-34 family in cancer and apoptosis." Cell Death Differ 17(2): 193-9.

Hershko, A. and A. Ciechanover (1998). "The ubiquitin system." Annu Rev Biochem 67: 425-79.

Hobert, O. (2007). "miRNAs play a tune." Cell 131(1): 22-4.

Hock, J., L. Weinmann, et al. (2007). "Proteomic and functional analysis of Argonaute-containing mRNA-protein complexes in human cells." EMBO Rep 8(11): 1052-60.

Holcik, M. and N. Sonenberg (2005). "Translational control in stress and apoptosis." Nat Rev Mol Cell Biol 6(4): 318-27.

Huang da, W., B. T. Sherman, et al. (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nat Protoc 4(1): 44-57.

Iakoucheva, L. M., C. J. Brown, et al. (2002). "Intrinsic disorder in cell-signaling and cancer-associated proteins." J Mol Biol 323(3): 573-84.

Ingolia, N. T., S. Ghaemmaghami, et al. (2009). "Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling." Science 324(5924): 218-23.

Ishihama, Y., Y. Oda, et al. (2005). "Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein." Mol Cell Proteomics 4(9): 1265-72.

Ishihama, Y., J. Rappsilber, et al. (2002). "Microcolumns with self-assembled particle frits for proteomics." J Chromatogr A 979(1-2): 233-9.

Jackson, A. L., J. Burchard, et al. (2006). "Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity." RNA 12(7): 1179-87.

Karginov, F. V., C. Conaco, et al. (2007). "A biochemical approach to identifying microRNA targets." Proc Natl Acad Sci U S A 104(49): 19291-6.

Keene, J. D. (2007). "RNA regulons: coordination of post-transcriptional events." Nat Rev Genet 8(7): 533-43.

Khvorova, A., A. Reynolds, et al. (2003). "Functional siRNAs and miRNAs exhibit strand bias." Cell 115(2): 209-16.

Klipp, E. (2009). "Timing matters." FEBS Lett 583(24): 4013-8.

Komili, S. and P. A. Silver (2008). “Coupling and coordination in gene expression processes: a systems biology view.” Nat Rev Genet 9(1): 38-48.

Komurov, K. and M. White (2007). “Revealing static and dynamic modular architecture of the eukaryotic protein interaction network.” Mol Syst Biol 3: 110.

Kosik, K. S. (2006). “The neuronal microRNA system.” Nat Rev Neurosci 7(12): 911-20.

Kozak, M. (1991). “Structural features in eukaryotic mRNAs that modulate the initiation of translation.” J Biol Chem 266(30): 19867-70.

Kramer, G., R. R. Sprenger, et al. (2009). “Identification and quantitation of newly synthesized proteins in Escherichia coli by enrichment of azidohomoalanine-labeled peptides with diagonal chromatography.” Mol Cell Proteomics 8(7): 1599-611.

Krek, A., D. Grun, et al. (2005). “Combinatorial microRNA target predictions.” Nat Genet 37(5): 495-500.

Krijgsveld, J., R. F. Ketting, et al. (2003). “Metabolic labeling of C. elegans and D. melanogaster for quantitative proteomics.” Nat Biotechnol 21(8): 927-31.

Kruger, M., M. Moser, et al. (2008). “SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function.” Cell 134(2): 353-64.

Lagos-Quintana, M., R. Rauhut, et al. (2002). “Identification of tissue-specific microRNAs from mouse.” Curr Biol 12(9): 735-9.

Lal, A., F. Navarro, et al. (2009). “miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to “seedless” 3’UTR microRNA recognition elements.” Mol Cell 35(5): 610-25.

Landgraf, P., M. Rusu, et al. (2007). “A mammalian microRNA expression atlas based on small RNA library sequencing.” Cell 129(7): 1401-14.

Lee, R. C., R. L. Feinbaum, et al. (1993). “The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.” Cell 75(5): 843-54.

Liao, L., S. K. Park, et al. (2008). “Quantitative proteomic analysis of primary neurons reveals diverse changes in synaptic protein content in fmr1 knockout mice.” Proc Natl Acad Sci U S A 105(40): 15281-6.

Lim, L. P., N. C. Lau, et al. (2005). “Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs.” Nature 433(7027): 769-73.

Lin, B., J. T. White, et al. (2005). “Evidence for the presence of disease-perturbed networks in prostate cancer cells by genomic and proteomic analyses: a systems approach to disease.” Cancer Res 65(8): 3081-91.

Liu, H., R. G. Sadygov, et al. (2004). “A model for random sampling and estimation of relative protein abundance in shotgun proteomics.” Anal Chem 76(14): 4193-201.

Liu, N. and E. N. Olson “MicroRNA regulatory networks in cardiovascular development.” Dev Cell 18(4): 510-25.

Livak, K. J. and T. D. Schmittgen (2001). “Analysis of relative gene expression data using real-

time quantitative PCR and the 2(-Delta Delta C(T)) Method." Methods 25(4): 402-8.

Looso, M., T. Borchardt, et al. (2010). "Advanced identification of proteins in uncharacterized proteomes by pulsed in vivo stable isotope labeling-based mass spectrometry." Mol Cell Proteomics 9(6): 1157-66.

Lu, P., C. Vogel, et al. (2007). "Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation." Nat Biotechnol 25(1): 117-24.

Lunde, B. M., C. Moore, et al. (2007). "RNA-binding proteins: modular design for efficient function." Nat Rev Mol Cell Biol 8(6): 479-90.

Mallick, P., M. Schirle, et al. (2007). "Computational prediction of proteotypic peptides for quantitative proteomics." Nat Biotechnol 25(1): 125-31.

Malmstrom, J., M. Beck, et al. (2009). "Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans." Nature 460(7256): 762-5.

Mamane, Y., E. Petroulakis, et al. (2006). "mTOR, translation initiation and cancer." Oncogene 25(48): 6416-22.

Marcotte, E. M. (2003). "Assembling a jigsaw puzzle with 20,000 parts." Genome Biol 4(6): 323.

Marcotte, E. M. (2007). "How do shotgun proteomics algorithms identify proteins?" Nat Biotechnol 25(7): 755-7.

Marcus, J. S., W. F. Anderson, et al. (2006). "Microfluidic single-cell mRNA isolation and analysis." Anal Chem 78(9): 3084-9.

Mathonnet, G., M. R. Fabian, et al. (2007). "MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F." Science 317(5845): 1764-7.

Milner, E., E. Barnea, et al. (2006). "The turnover kinetics of major histocompatibility complex peptides of human cancer cells." Mol Cell Proteomics 5(2): 357-65.

Minezaki, Y., K. Homma, et al. (2006). "Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation." J Mol Biol 359(4): 1137-49.

Mishima, Y., A. J. Giraldez, et al. (2006). "Differential regulation of germline mRNAs in soma and germ cells by zebrafish miR-430." Curr Biol 16(21): 2135-42.

Moldave, K. (1985). "Eukaryotic protein synthesis." Annu Rev Biochem 54: 1109-49.

Morris, D. R. and A. P. Geballe (2000). "Upstream open reading frames as regulators of mRNA translation." Mol Cell Biol 20(23): 8635-42.

Mortazavi, A., B. A. Williams, et al. (2008). "Mapping and quantifying mammalian transcriptomes by RNA-Seq." Nat Methods 5(7): 621-8.

Mosley, A. L., L. Florens, et al. (2009). "A label free quantitative proteomic analysis of the Saccharomyces cerevisiae nucleus." J Proteomics 72(1): 110-20.

Nagalakshmi, U., K. Waern, et al. (2010). "RNA-Seq: a method for comprehensive transcriptome analysis." Curr Protoc Mol Biol Chapter 4: Unit 4 11 1-13.

Nie, L., G. Wu, et al. (2006). “Correlation between mRNA and protein abundance in Desulfovibrio vulgaris: a multiple regression to identify sources of variations.” Biochem Biophys Res Commun 339(2): 603-10.

Nie, L., G. Wu, et al. (2006). “Correlation of mRNA expression and protein abundance affected by multiple sequence features related to translational efficiency in Desulfovibrio vulgaris: a quantitative analysis.” Genetics 174(4): 2229-43.

Ning, K., D. Fermin, et al. (2010). “Computational analysis of unassigned high-quality MS/MS spectra in proteomic data sets.” Proteomics 10(14): 2712-8.

Nottrott, S., M. J. Simard, et al. (2006). “Human let-7a miRNA blocks protein production on actively translating polyribosomes.” Nat Struct Mol Biol 13(12): 1108-14.

Ong, S. E., B. Blagoev, et al. (2002). “Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.” Mol Cell Proteomics 1(5): 376-86.

Ong, S. E. and M. Mann (2005). “Mass spectrometry-based proteomics turns quantitative.” Nat Chem Biol 1(5): 252-62.

Ong, S. E. and M. Mann (2006). “A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC).” Nat Protoc 1(6): 2650-60.

Orntoft, T. F., T. Thykjaer, et al. (2002). “Genome-wide study of gene copy numbers, transcripts, and protein levels in pairs of non-invasive and invasive human transitional cell carcinomas.” Mol Cell Proteomics 1(1): 37-45.

PaÅ¡a-ToliÄ‡, L., P. K. Jensen, et al. (1999). “High Throughput Proteome-Wide Precision Measurements of Protein Expression Using Mass Spectrometry.” Journal of the American Chemical Society 121(34): 7949-7950.

Parker, R. and U. Sheth (2007). “P bodies and the control of mRNA translation and degradation.” Mol Cell 25(5): 635-46.

Paulsson, J. (2005). “Models of stochastic gene expression.” Physics of Life Reviews 2(2): 157-175.

Pratt, J. M., J. Petty, et al. (2002). “Dynamics of protein turnover, a missing dimension in proteomics.” Mol Cell Proteomics 1(8): 579-91.

Rajewsky, N. (2006). “microRNA target predictions in animals.” Nat Genet 38 Suppl: S8-13.

Rajewsky, N. and N. D. Socci (2004). “Computational identification of microRNA targets.” Dev Biol 267(2): 529-35.

Rappsilber, J., Y. Ishihama, et al. (2003). “Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics.” Anal Chem 75(3): 663-70.

Rappsilber, J., M. Mann, et al. (2007). “Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips.” Nat Protoc 2(8): 1896-906.

Reeds, P. J. and T. A. Davis (1999). “Of flux and flooding: the advantages and problems of different isotopic methods for quantifying protein turnover in vivo: I. Methods based on the dilution of a tracer.” Curr Opin Clin Nutr Metab Care 2(1): 23-8.

Reinhart, B. J., F. J. Slack, et al. (2000). “The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans.” Nature 403(6772): 901-6.

Rogers, S., R. Wells, et al. (1986). “Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis.” Science 234(4774): 364-8.

Rosenfeld, N., J. W. Young, et al. (2005). “Gene regulation at the single-cell level.” Science 307(5717): 1962-5.

Ross, P. L., Y. N. Huang, et al. (2004). “Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.” Mol Cell Proteomics 3(12): 1154-69.

Scheper, G. C., M. S. van der Knaap, et al. (2007). “Translation matters: protein synthesis defects in inherited disease.” Nat Rev Genet 8(9): 711-23.

Schuster, S. C. (2008). “Next-generation sequencing transforms today’s biology.” Nat Methods 5(1): 16-8.

Schwanhausser, B., M. Gossen, et al. (2009). “Global analysis of cellular protein translation by pulsed SILAC.” Proteomics 9(1): 205-9.

Scigelova, M. and A. Makarov (2006). “Orbitrap mass analyzer - overview and applications in proteomics.” Proteomics 6 Suppl 2: 16-21.

Selbach, M., B. Schwanhausser, et al. (2008). “Widespread changes in protein synthesis induced by microRNAs.” Nature 455(7209): 58-63.

Sharova, L. V., A. A. Sharov, et al. (2009). “Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells.” DNA Res 16(1): 45-58.

Shaw, G. and R. Kamen (1986). “A conserved AU sequence from the 3’ untranslated region of GM-CSF mRNA mediates selective mRNA degradation.” Cell 46(5): 659-67.

Shevchenko, A., H. Tomas, et al. (2006). “In-gel digestion for mass spectrometric characterization of proteins and proteomes.” Nat Protoc 1(6): 2856-60.

Silva, J. C., M. V. Gorenstein, et al. (2006). “Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition.” Mol Cell Proteomics 5(1): 144-56.

Sood, P., A. Krek, et al. (2006). “Cell-type-specific signatures of microRNAs on target mRNA expression.” Proc Natl Acad Sci U S A 103(8): 2746-51.

Stark, A., J. Brennecke, et al. (2003). “Identification of Drosophila MicroRNA targets.” PLoS Biol 1(3): E60.

States, D. J., G. S. Omenn, et al. (2006). “Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study.” Nat Biotechnol 24(3): 333-8.

Steen, H. and M. Mann (2004). “The ABC’s (and XYZ’s) of peptide sequencing.” Nat Rev Mol Cell Biol 5(9): 699-711.

Steen, H. and A. Pandey (2002). “Proteomics goes quantitative: measuring protein abundance.” Trends Biotechnol 20(9): 361-4.

Stefani, G. and F. J. Slack (2008). “Small non-coding RNAs in animal development.” Nat Rev Mol Cell Biol 9(3): 219-30.

Stenvang, J. and S. Kauppinen (2008). “MicroRNAs as targets for antisense-based therapeutics.” Expert Opin Biol Ther 8(1): 59-81.

Taniguchi, Y., P. J. Choi, et al. (2010). “Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells.” Science 329(5991): 533-8.

Thermann, R. and M. W. Hentze (2007). “Drosophila miR2 induces pseudo-polysomes and inhibits translation initiation.” Nature 447(7146): 875-8.

Tian, Q., S. B. Stepaniants, et al. (2004). “Integrated genomic and proteomic analyses of gene expression in Mammalian cells.” Mol Cell Proteomics 3(10): 960-9.

Tompa, P. (2002). “Intrinsically unstructured proteins.” Trends Biochem Sci 27(10): 527-33.

Tompa, P., J. Prilusky, et al. (2008). “Structural disorder serves as a weak signal for intracellular protein degradation.” Proteins 71(2): 903-9.

Unwin, R. D., D. L. Smith, et al. (2006). “Quantitative proteomics reveals posttranslational control as a regulatory factor in primary hematopoietic stem cells.” Blood 107(12): 4687-94.

Valencia-Sanchez, M. A., J. Liu, et al. (2006). “Control of translation and mRNA degradation by miRNAs and siRNAs.” Genes Dev 20(5): 515-24.

van Rooij, E. and E. N. Olson (2007). “MicroRNAs: powerful new regulators of heart disease and provocative therapeutic targets.” J Clin Invest 117(9): 2369-76.

Vasudevan, S., Y. Tong, et al. (2007). “Switching from repression to activation: microRNAs can up-regulate translation.” Science 318(5858): 1931-4.

Voet, D. V. a. J. G. (1995). Biochemistry. New York, John Wiley and Sons.

Vogel, C., S. Abreu Rde, et al. (2010). “Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line.” Mol Syst Biol 6: 400.

Wahlestedt, C., P. Salmi, et al. (2000). “Potent and nontoxic antisense oligonucleotides containing locked nucleic acids.” Proc Natl Acad Sci U S A 97(10): 5633-8.

Wakiyama, M., K. Takimoto, et al. (2007). “Let-7 microRNA-mediated mRNA deadenylation and translational repression in a mammalian cell-free system.” Genes Dev 21(15): 1857-62.

Wang, X., J. McLachlan, et al. (2002). “Modular recognition of RNA by a human pumilio-homology domain.” Cell 110(4): 501-12.

Wang, Z., M. Gerstein, et al. (2009). “RNA-Seq: a revolutionary tool for transcriptomics.” Nat Rev Genet 10(1): 57-63.

Ward, J. J., J. S. Sodhi, et al. (2004). “Prediction and functional analysis of native disorder in proteins from the three kingdoms of life.” J Mol Biol 337(3): 635-45.

Wightman, B., I. Ha, et al. (1993). “Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans.” Cell 75(5): 855-62.

Wilkins, M. (2009). “Proteomics data mining.” Expert Rev Proteomics 6(6): 599-603.

Wilusz, C. J., M. Wormington, et al. (2001). “The cap-to-tail guide to mRNA turnover.” Nat Rev Mol Cell Biol 2(4): 237-46.

Wu, C. C., M. J. MacCoss, et al. (2004). “Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis.” Anal Chem 76(17): 4951-9.

Wu, L., J. Fan, et al. (2006). “MicroRNAs direct rapid deadenylation of mRNA.” Proc Natl Acad Sci U S A 103(11): 4034-9.

Yao, X., A. Freas, et al. (2001). “Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus.” Anal Chem 73(13): 2836-42.

VII Supplementary information

VII.1.Publication list

Published articles:

- **Schwanhaeusser B,** Gossen M, Dittmar G, Selbach M. (2008). *Global analysis of cellular protein translation by pulsed SILAC*. Proteomics. 2009 Jan; 9(1):205-9.
- Selbach M, **Schwanhaeusser B**, Thierfelder N, Fang Z, Khanin R, Rajewsky N. (2008). *Widespread changes in protein synthesis induced by microRNAs*. Nature 455(7209): 58-63.
- **Schwanhaeusser B**., Selbach M., "Fachbeitrag" for the BioForum entitled "*Globale Anayse der zellulären Proteintranslation mit pulsed SILAC*". The article has been published in the May 2009 issue.

Manuscripts in preparation or under editorial consideration:

- **Schwanhaeusser B**, Cox J, Kirchner M, Mann M, Selbach M. *Simple and accurate quantification of absolute protein abundance* (manuscript in preparation).
- **Schwanhaeusser B,** Busse D, Li N, Dittmar G, Wolf J, Chen W, Selbach M. *Genome-wide parallel quantification of mRNA and protein levels and turnover in mammalian cells* (manuscript submitted).

Co-authored manuscripts in preparation:

- Lebedeva S, Jens M, **Schwanhaeusser B**, Feldkamp M, Langnick C, Chen W, Selbach M, Landthaler M, Rajewsky N. *Genome-wide regulatory interactions of the RNA-binding protein HuR* (ELAVL1).
- Baltz A, **Schwanhaeusser B**, Chen W, Selbach M, Landthaler M. *Biochemical characterization of the protein-mRNA Interactome of HEK293 cells.*

Berlin, den

_____

Björn Schwanhäußer

VII.2. Awards

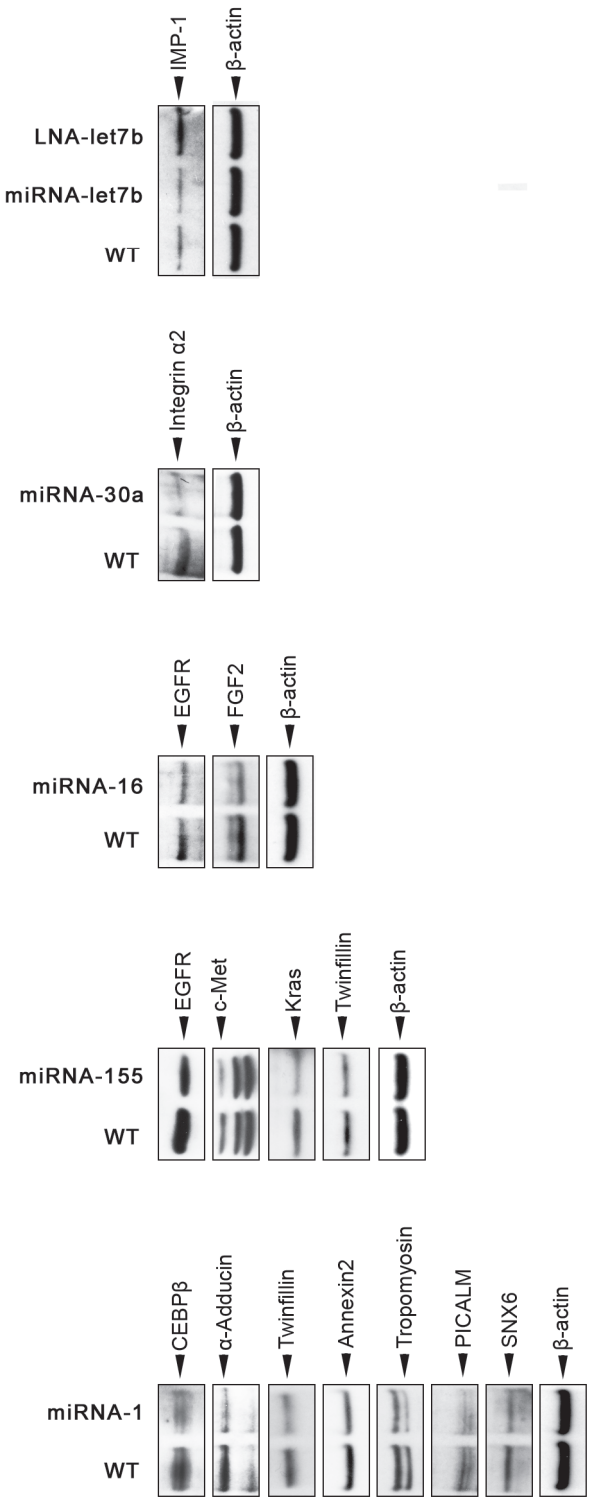
- For the work on the impact of microRNAs on gene expression, the authors of the paper “Widespread changes in protein synthesis induced by microRNAs” awarded the Forschungspreis der Deutschen Gesellschaft für Gentherapie” founded by the Coley Pharmaceutical Company.

VII.3. Posters and talks

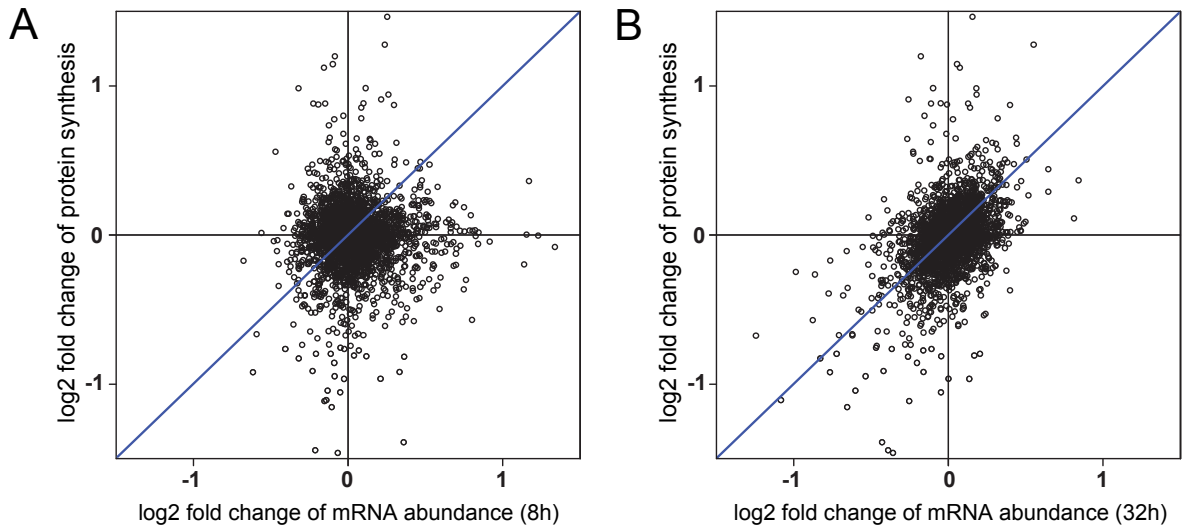
- **Keystone Symposia 2008, Coeur d´Alene, USA.** “Global analysis of cellular protein translation by pulsed SILAC” (Poster presentation).
- **Berlin Summer Meeting 2008, Berlin, Germany.** “Wide-spread changes in protein synthesis induced by microRNAs” (Oral presentation).
- **8th annual Human Proteome Congress (HUPO2009), Toronto, Canada. “**System-wide in depth analysis of protein and transcript half-lives” (Poster presentation).
- **2nd annual meeting of the National Genome Research Network (NGFN 2009), Berlin, Germany.** “Genome-wide analysis of mRNA and protein half-lives reveals dynamic properties of mammalian gene expression” (Oral presentation).

VII.4.Supplementary data

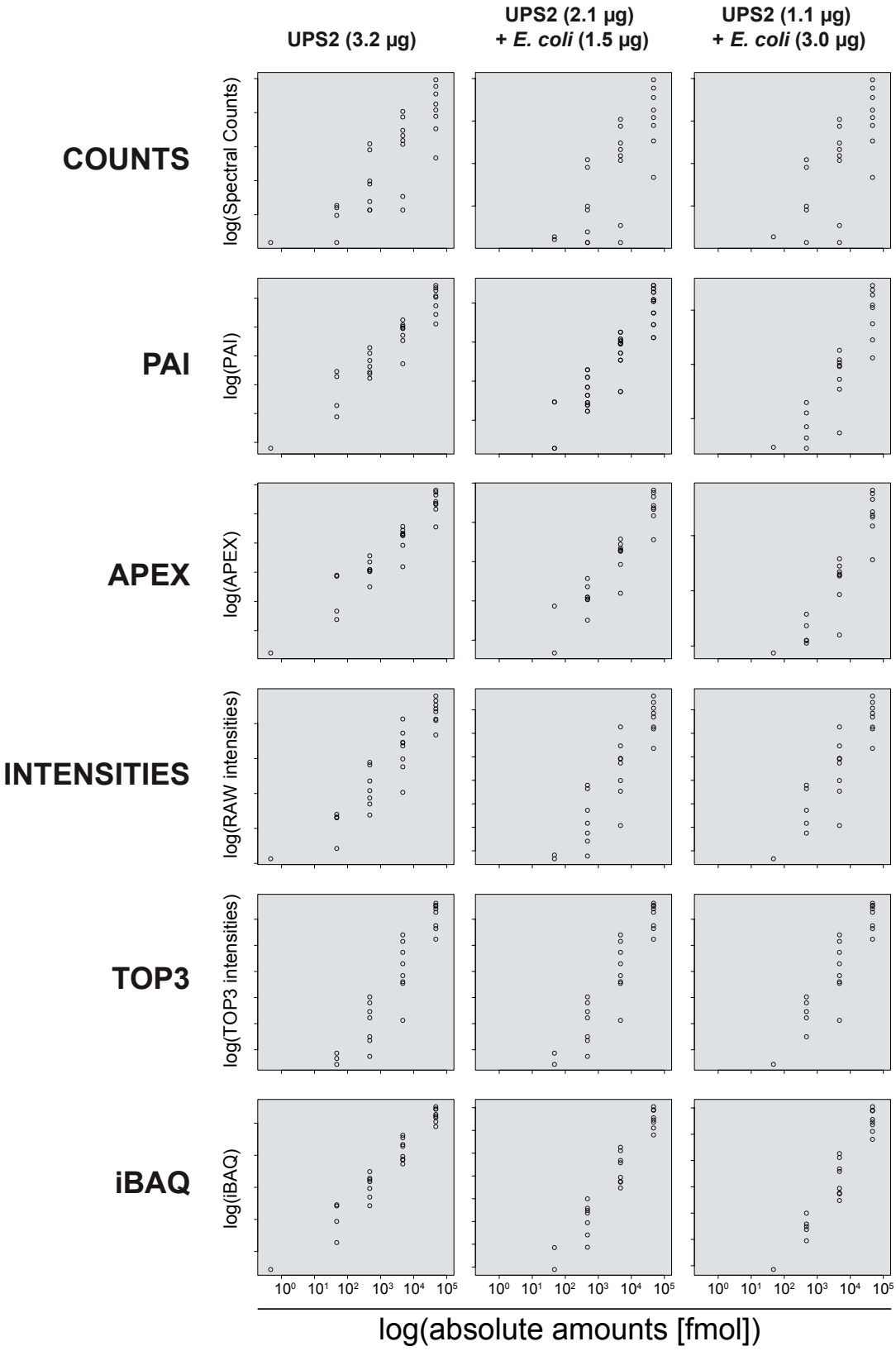
VII.4.1. Supplementary figures



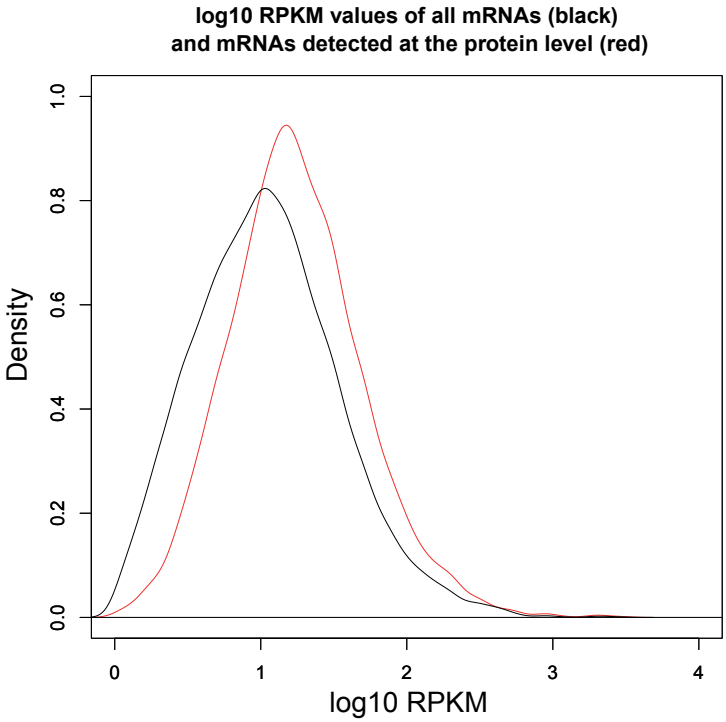
Supplementary Fig. VII.1 | Verification of targets identified by pSILAC with Western blotting. HeLa cells were transfected with respective miRNAs and LNA. WT (Wild-type) represents "water-control" (cf. Material & Methods II.8). Samples were analysed by Western blotting with antibodies against c-Met, EGFR, α-Adducin, CEBPβ, Twinfilin, Annexin2, Tropomyosin, PICALM, SNX6, Kras, FGF2, Integrin-α2 and IMP-1. β-Actin served as a loading control.



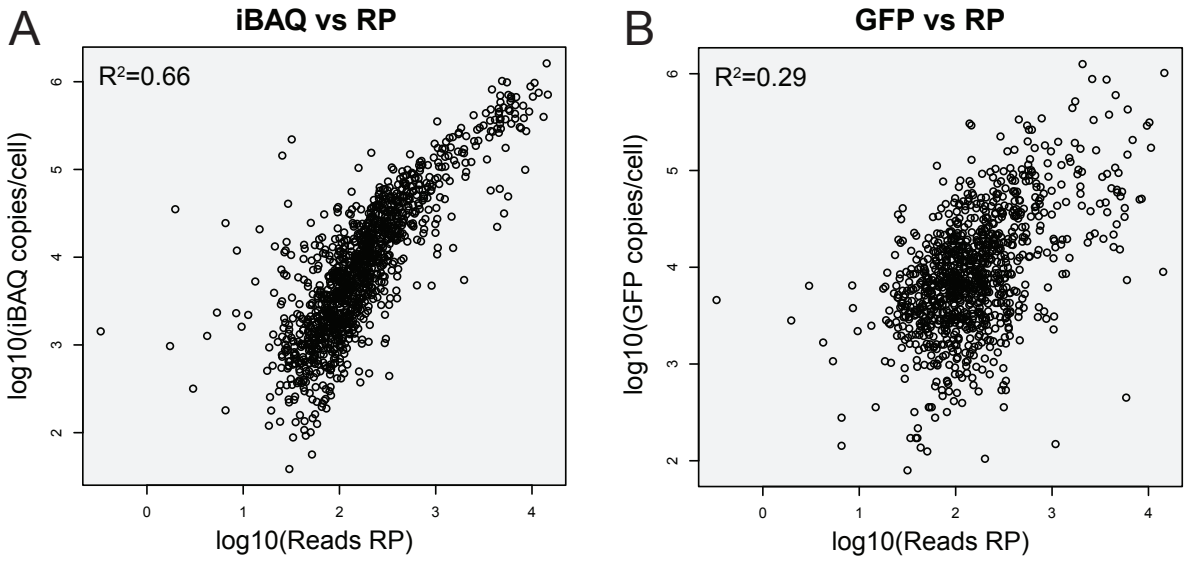
Supplementary Fig. VII.2 | miRNAs inhibit translation on a genome-wide scale. (A) Changes in protein production between 8 and 32 h after miR-30a transfection with mRNA fold-changes at 8 h reveal poor overall correlation. (B) mRNA levels at 32 h correlate remarkably well with changes in protein synthesis.



Supplementary Fig. VII.3 | Correlation between absolute amounts and values calculated by various approaches of absolute quantification. Dots designate individual proteins of the standard (UPS2). Each plot is based on three technical replicates.



Supplementary Fig. VII.4 | Kernel density estimates of mRNA levels (given as RPKM values) for all genes (black) and genes detected on protein level (red). The slight shift indicates a mild detection bias against lowly expressed proteins. All subsequent calculations were therefore done on the set of genes detected on both the protein and mRNA level.



Supplementary Fig. VII.5 | Comparison of yeast protein copy numbers obtained by either iBAQ based absolute quantification (A) or tag-western-blot/GFP-fusion proteins (Ghaemmaghami *et al.,* 2003) (B) with ribosome profiling (RP) data (Ingolia *et al.*, 2009).

VII.4.2. Supplementary tables

Supplementary Table VII.1: Pulsed SILAC ratios of all quantified peptides of luciferase after doxycycline treatment for different length of time

Time	Peptide-Sequence	Mass	Mass Error [ppm]	Log2 fold change luciferase peptides	Log2 fold change luciferase (median of peptide ratios)
1 h	EIVDYVASQVTTAK	1522.79	-0.27	3.24	5.99
	GGVVFVDEVPK	1144.61	0.06	4.33	
	IQSALLVPTLFSFFAK	1781.01	0.19	7.19;5.77;6.22	
	VVDLDTGK	845.45	-0.70	6.94	
2 h	EIVDYVASQVTTAK	1522.79	-0.03	5.15;3.78	4.65
	FEEELFLR	1081.54	-0.40	5.34	
	GGVVFVDEVPK	1144.61	-0.30	5.34	
	GPMIMSGYVNNPEATNALIDK	2234.07	-0.59	2.86;2.20	
	IIIMDSK	818.46	0.69	5.13	
	IQSALLVPTLFSFFAK	1781.01	0.55	5.05;4.25	
	KGPAPFYPLEDGTAGEQLHK	2154.07	-0.49	2.88	
	STLIDKYDLSNLHEIASGGAPLSK	2528.31	0.13	3.19	
	TIALIMNSSGSTGLPK	1588.85	-0.38	5.11	
4 h	EIVDYVASQVTTAK	1522.79	-0.60	1.82	2.00
	FEEELFLR	1081.54	-0.58	2.51	
	GGVVFVDEVPK	1144.61	-0.11	1.66	
	IQSALLVPTLFSFFAK	1781.01	0.24	2.00;2.01;1.97;1.71;2.40	
	STLIDKYDLSNLHEIASGGAPLSK	2528.31	0.29	2.11	
	TIALIMNSSGSTGLPK	1588.85	-0.34	2.07	
	VVDLDTGK	845.45	0.33	1.97	
6 h	EIVDYVASQVTTAK	1522.79	0.14	0.94;0.47;1.07;1.15	1.05
	FEEELFLR	1081.54	-0.16	1.13;0.84	
	GGVVFVDEVPK	1144.61	-0.25	0.88	
	IIIMDSK	818.46	0.93	1.12	
	IQSALLVPTLFSFFAK	1781.01	0.15	1.05;1.02;0.95;1.25;1.20;1.25	
	KGPAPFYPLEDGTAGEQLHK	2154.07	-0.55	1.02;0.66	
	TIALIMNSSGSTGLPK	1588.85	-0.71	1.25	
8 h	EIVDYVASQVTTAK	1522.79	-0.41	0.25;0.41	0.38
	FEEELFLR	1081.54	-0.13	0.58	
	GGVVFVDEVPK	1144.61	0.08	0.30;0.30	
	GPMIMSGYVNNPEATNALIDK	2234.07	-0.57	0.43;0.58	
	GVALPHR	748.43	-1.33	0.32	
	IIIMDSK	818.46	0.30	0.37	
	IQSALLVPTLFSFFAK	1781.01	0.15	0.35;0.50;0.44;0.28;0.70;0.14	
	KGPAPFYPLEDGTAGEQLHK	2154.07	-0.61	0.33;0.14	
	STLIDKYDLSNLHEIASGGAPLSK	2528.31	0.23	0.58	
	TIALIMNSSGSTGLPK	1588.85	-0.29	0.43	
	VVDLDTGK	845.45	-0.11	0.40	
10 h	EIVDYVASQVTTAK	1522.79	-0.02	-0.15;-0.16;0.00	-0.09
	FEEELFLR	1081.54	-0.50	-0.08	
	GGVVFVDEVPK	1144.61	-0.83	-0.18	
	IQSALLVPTLFSFFAK	1781.01	-0.26	0.16;0.05;-0.10	
	KGPAPFYPLEDGTAGEQLHK	2154.07	-1.28	-0.39;0.05	
	QGYGLTETTSAILITPEGDDKPGAVGK	2717.38	0.59	-0.19	
	TIALIMNSSGSTGLPK	1588.85	-0.52	0.03	
	VVDLDTGK	845.45	0.01	0.00	
	YDLSNLHEIASGGAPLSK	1870.94	-0.30	-0.27	

Supplementary Table VII.2: Pulsed SILAC ratios of all quantified peptides of ferritin H and L chain and the transferrin receptor after iron treatment of Hela cells

Protein	Peptide-Sequence	Mass	Mass Error [ppm]	Log2 fold change peptides	Log2 fold change proteins (median of peptide ratios)
Ferritin H (IPI00554521)	AIKELGDHVTNLRK	1592.90	-0.30	4.13	3.74
	ELGDHVTNLR	1152.59	-0.06	2.88;3.38;3.58;3.76;3.77;4.56	
	IFLQDIK	875.51	0.01	-0.41;2.42;2.44;3.34;3.50;3.63;3.74;4.13	
	KMGAPESGLAEYLFDK	1754.85	0.44	3.23	
	MGAPESGLAEYLFDK	1626.76	0.03	3.07;3.58;3.74;4.19;4.35	
	MGAPESGLAEYLFDKHTLGDSDNES	2682.18	-0.07	3.98;4.36	
	NVNQSLLELHK	1293.70	0.00	3.48;3.52;3.54;3.62;3.71;3.73;3.77;3.95;4.03	
	QNYHQDSEAAINR	1544.70	-0.01	4.01;4.08	
	YFLHQSHEER	1344.62	-0.07	3.87;4.06	
Ferritin L (IPI00738499)	KLNQALLDLHALGSAR	1718.98	0.03	3.36;3.66;3.76	3.68
	LGGPEAGLGEYLFER	1606.80	0.03	2.85;3.65;3.67;3.80;3.87;3.88	
	LNQALLDLHALGSAR	1590.88	-0.01	3.43;3.65;3.68;3.82;3.89	
	MGDHLTNLHR	1192.58	0.12	3.44;3.88	
Transferrin Receptor Protein (IPI00022462)	VNILEVASGAVLR	1339.78	-0.09	-0.99;-0.20;-0.00	-1.62
	AFTYINLDK	1083.56	-0.04	1.13	
	AVLGTSNFK	935.51	-0.02	-1.94;-1.50;-1.44	
	DENLALYVENQFR	1609.77	-0.09	-1.62;-1.38;0.43	
	DSAQNSVIIVDK	1287.67	-0.15	-2.68;-2.55;-1.30	
	EAGSQKDENLALYVENQFR	2210.06	0.05	-1.92;-1.63	
	GFVEPDHYVVVGAQR	1671.84	-0.06	-1.75;-1.58;-1.51	
	HPVTGQFLYQDSNWASK	1976.94	-0.25	-2.71;-2.13;-1.60;-1.05	
	ILNIFGVIK	1015.64	-0.01	-1.85;-1.83;-1.62;-1.58;-1.56;-1.55	
	LAQMFSDMVLK	1281.65	0.01	-1.74;-1.23;-1.15;0.07	
	LAVDEEENADNNTK	1560.69	0.02	-2.11;-1.04	
	LDSTDFTGTIK	1196.59	0.00	-2.16;-1.65;-1.60;-0.83	
	LLNENSYVPR	1203.62	0.05	-2.32;-2.11;-1.96;-1.21;-0.82	
	LTHDVELNLDYER	1615.78	-0.10	-2.30;-1.65;-1.64;-1.37	
	LTTDFGNAEK	1094.52	0.00	-1.46;-1.09	
	LVHANFGTK	985.53	-0.10	-1.75;-1.21	
	LVYLVENPGGYVAYSK	1770.92	0.03	-1.64	
	QVDGDNSHVEMK	1357.59	0.04	-1.44	
	RLYWDDLKR	1263.67	-0.06	-1.85;-0.63	
	SAFSNLFGGEPLSYTR	1744.84	-0.03	-3.08;-1.79;-1.79;-1.77;-1.26;-0.19	
	SGVGTALLLK	957.59	0.00	-2.21;-1.69;-1.56	
	SSGLPNIPVQTISR	1467.80	-0.08	-1.82;-1.75;-1.71;-1.25	
	VEYHFLSPYVSPK	1564.79	0.00	-2.07;-1.77;-1.76;-1.41;-1.27;-0.86	
	VSASPLLYTLIEK	1432.82	0.01	-1.81;-1.66;-1.50	

Supplementary Table VII.3: Pulsed SILAC ratios of all quantified peptides of five arbitrarily selected proteins that remain unaffected (log2 fold change of ~ 0) upon iron treatment of Hela cells

Protein	Peptide-Sequence	Mass	Mass Error [ppm]	Log2 fold change peptides	Log2 fold change proteins (median of peptide ratios)
Serine/threonine-protein kinase PAK 2 (IPI00419979)	KNPQAVLDVLK	1223.72	-0.89	0.32	**0.04**
	LAKPLSSLTPLIMAAK	1652.99	0.01	0.00	
	LLQTSNITK	1016.59	-0.29	-0.06	
	SDNVLLGMEGSVK	1347.67	0.50	-0.09	
	STMVGTPYWMAPEVVTR	1923.92	-0.09	0.08	
	SVIDPVPAPVGDSHVDGAAK	1929.98	0.01	0.64	
Eukaryotic translation initiation factor 5 (IPI00022648)	ALNRPPTYPTK	1256.69	-0.29	0.03	**0.03**
	AMGPLVLTEVLFNEK	1659.89	0.08	-0.56;0.07;0.18	
	SDNKDDDIDIDAI	1447.63	0.02	-0.77;0.24	
	TVIVNMVDVAK	1187.66	0.00	-0.21;0.03	
	VLTLSDDLER	1159.61	0.08	-0.39;-0.21;-0.04	
	VNILFDFVK	1093.62	0.05	-0.17;0.19;0.34;0.47	
Actin-related protein 3 (IPI00028091)	AEPEDHYFLLTEPPLNTPENR	2481.18	0.00	-0.25;-0.19;-0.16	**0.01**
	DITYFIQQLLR	1408.77	-0.02	-0.42;0.10;0.15;0.32;0.46;2.11	
	EFSIDVGYER	1213.56	-0.03	0.48	
	FMEQVIFK	1040.54	-0.01	-0.32;-0.00;0.08	
	GVDDLDFFIGDEAIEKPTYATK	2443.18	0.19	-0.27;-0.07;0.23	
	HGIVEDWDLMER	1498.69	-0.02	0.02;0.05;0.49	
	LKPKPIDVQVITHHMQR	2039.15	-0.14	-3.15;-0.90;-0.54;0.37	
	NIVLSGGSTMFR	1280.65	-0.05	-0.40;0.09	
	YSYVCPDLVK	1242.60	-0.13	-0.19	
Ras-related protein Rab-21 (IPI00007755)	AAAGGGGGGAAAAGR	1070.52	-0.41	-0.71	**-0.01**
	DSNGAILVYDITDEDSFQK	2128.98	0.00	0.18	
	FHALGPIYYR	1235.65	0.02	-0.48;-0.09;-0.08;-0.01	
	HVSIQEAESYAESVGAK	1803.86	0.02	-0.10;0.08;0.35	
	MIETAQVDER	1190.56	-0.02	0.15;0.27	
	RVNLAIWDTAGQER	1627.84	-0.02	-0.01	
	VNLAIWDTAGQER	1471.74	0.00	-0.14;-0.06;0.37	
GMP synthase (IPI00029079)	DEPDWESLIFLAR	1589.77	0.10	0.05	**-0.01**
	ELFVQSEIFPLETPAFAIK	2178.16	0.10	-0.01	
	ELGLPEELVSR	1240.67	0.03	-2.41;0.10;0.27;0.47	
	HPFPGPGLAIR	1160.65	0.11	-0.46	
	LMQITSLHSLNAFLLPIK	2038.16	0.29	-0.95	
	SGNIVAGIANESK	1258.65	-0.12	1.58	
	VIEPLKDFHKDEVR	1723.93	-0.16	-0.25	
	VVYIFGPPVK	1117.65	0.10	-2.26;0.19	
	VVYIFGPPVKEPPTDVTPTFLTTGVLSTLR	3243.78	-0.01	-0.08	

VIII Acknowledgements

The good thing is that being excited is always part of a scientist´s life just because most experiments are expeditions into uncharted areas. The bad thing is that too many (haphazard) journeys into the unknown can drive you insane ("expedition choler", Lorenz, 1971). Hence, the acknowledgement part is where I would like to thank all the forces that helped me in mastering my years as a PhD student. First and foremost, I would like to express my warmest thanks to my supervisor Dr. Matthias Selbach. His enthusiasm and passion for science, his sincere guidance and contagious optimism were more than I could ever expect. I am much obliged to Prof. Thomas Sommer for accepting the official supervision at the Humboldt University. With his untiringly kindness he always had a sympathetic ear for whatever concerns I had. I am grateful to Prof. Andreas Herrmann for his interest in my work which has led to an exciting collaboration. I thank Prof. Nikolaus Rajewsky for joining efforts in exciting collaborations and for trenchant questions that relentlessly revealed the Achilles´ heel of yet every project -what made them undeniably better ones. I also thank Jennifer Stewart for sharing her well-dosed caustic humour what is probably the only way to master her position. I am also greatly indebted to all my dear lab members Flo, Jimmy, Fabian, Christian, Livey, This, Marie and Murphy. Apart from outstanding scientific input (youtube) and your altruistic help in the lab (music), sharing life in our spacious but cosy and well-resourced (ac) office during summer and winter times with you was real fun. I thank Christian for taking care about our children of sorrow Ernie, Bert and Tiffy. I´ll miss the smell and sound of machines running on promising samples. With their somewhat erratic way of operating they always found a way to annihilate the most important last sample just to demonstrate their power and our dependency. I am much obliged to my collaborators Nadine Thierfelder, Zhou Fang and Raya Khanin for the work on microRNAs, Alexander Baltz and Markus Landthaler for protein-mRNA interactome studies and Svetlana Lebedeva for grappling with HuR. Special thanks go to Dorothea Busse, Jana Wolf, Gunnar Dittmar, Wei Chen and Na Li for endless hours discussing and puzzling about steady-state conditions. You brought me much closer to ordinary differential equations than what is commonly considered to be healthy. I also have to thank the MDC security staff for not firing me right after the yeast bomb accidently went off. It is an honour for me to thank Olli and Carsten for countless nights in Mini Bar, West Germany, Cake, Privatclub etc. as an excellent means of recreation. Next days hangover was always highly motivating. I owe deepest thanks to my family for unconditional support, persistent confidence in me and encouragement to pursue my not always easy to communicate interests. And of course, with more than words can say, my most sincere thanks deserves Bianca. Her unfailing support, affectionate humour and character gave me strength when I had none. I feel blessed for that beautiful collision and I have highest respect for what you went through in the last nine months. This is not the end, it is just about to begin!

IX Selbständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Hilfsmittel angefertigt habe.

Berlin, den 28.10.2010

_____

Björn Schwanhäußer